SPECIAL ISSUE

ETHICAL AI • OTHER EARTHS • FUSION • ANIMAL TESTING
GREEN ENERGY • POPULATION TRENDS • GALAXY ORIGINS

AMERICAN
Scientist

July–August 2023

www.americanscientist.org

How Real Are Models?

Simulations can address our biggest questions—if we are honest about their power and their limitations.

SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY

AMERICAN Scientist

Volume 111 • Number 4 • July–August 2023

CONTENTS

SPECIAL ISSUE: Scientific Modeling

ON THE COVER
Scientific models can help researchers explore large and complex topics, from computer simulations of galaxy formation and climate change to mouse models of human disease. However, scientists must also strive to ensure that such models are accurate, useful, and fair. (Cover illustration by Michael Morgenstern.)

From the Editors

Modeling Well

Modeling is a technique that cuts across scientific fields and uses many different technologies. A mathematical model can be used to study climate, or galaxy formation (pages 222–225), or cancer. An animal model can reveal new insights about a disease (pages 213–215). Models sometimes are employed to interpret processes involving subatomic particles, such as neutrino detection (pages 226–231) or nuclear fusion (pages 211–212). Models can be used to examine the past, such as the process of human evolution (pages 235–236), or they can be used to try to predict the future, such as in global demographics (pages 200–203).

Although models are ubiquitous in science, they are not a magic box, and they are only as good as we make them. Properly conceptualizing them is an art as well as a science, and their limits of applicability have to be carefully considered. In addition, scientists have to be vigilant in creating models that don't inadvertently introduce bias (pages 204–207), especially when artificial intelligence is employed in attempts to cope with the mountains of data that a model can use and produce.

In this special single-topic issue, we take a step back and give the big picture of what models can and can't do well. In "Approximating Reality" (pages 198–199), Kevin Heng of Ludwig Maximilian University in Germany kicks off the discussion with his reflections on what it means to understand a concept in the natural world, and how that definition of understanding can affect how a model of a phenomenon is built. Producing the most complex model, Heng notes, may not always be the goal. As he says: "We simulate in order to mimic as much of an observed phenomenon as possible. But we achieve understanding by using idealized models to capture the essence of the phenomenon."

Heng says that AI could be a boon to modeling, because automating the process of designing, writing, and interpreting models could help scientists figure out direct causation of effects more efficiently, allowing them to spend more time formulating deep, creative questions for the models to answer. But in "Bias Optimizers" (pages 204–207), Damien Patrick Williams of the University of North Carolina at Charlotte focuses more on the creation of AI and the inherent biases that it can amplify, making it difficult to root out even unintentional sources of problematic results.

There are many cases in which data selection can affect model performance. In "Building Better Growth Curves" (pages 216–221), William E. Bennett, Jr of Indiana University School of Medicine discusses the history of infant growth projections and why they can often lead to false positives when diagnosing children with what's called *failure to thrive*. But he also explains that efforts to personalize growth curves can raise serious concerns about data privacy and implicit bias.

Anne Goujon of the International Institute for Applied Systems Analysis in Austria discusses how historical demographic data can help countries plan for the future, in "Predicting an Aging and Changing World" (pages 200–203). Goujon describes models of Niger's future population that are based on whether or not funding is increased now for education, and explains how these different models can help policymakers decide where to use their limited resources to achieve the most prosperous outcomes for their countries.

Readers can find lots more in this issue, and additional online content is listed on page 196. Join us on social media to share your own experiences with modeling.

—Fenella Saunders (@FenellaSaunders)

AMERICAN **Scientist**

www.americanscientist.org

VOLUME 111, NUMBER 4

American Scientist gratefully acknowledges support for "Engineering" through the Leroy Record Fund.

ADDITIONAL DIGITAL CONTENT

Collection on Scientific Modeling

Using simulations to solve scientific problems is nothing new—check out the Engineering column (pages 232–234) to learn about Galileo's creative thinking—and neither is *American Scientist*'s coverage of the topic. The editors have collected pieces from our archives that delve into the many ways that scientists have used models.

How Reliable Are Models?

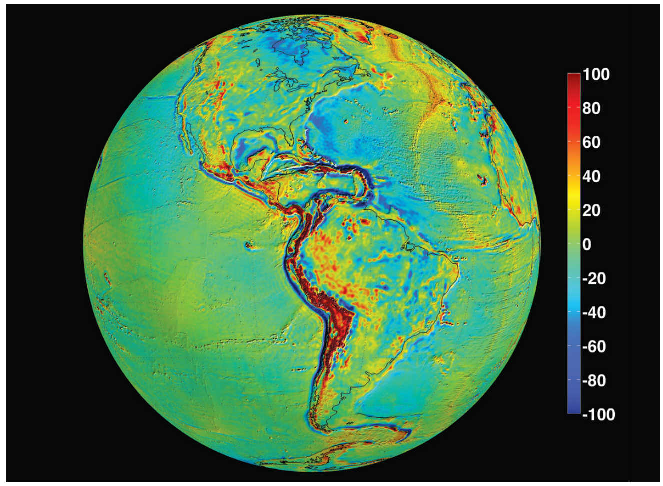
Mathematical descriptions and simulations help scientists forecast events and recommend actions, but it can be difficult to determine whether those predictions are trustworthy. For a primer on the topic, check out "In Models We Trust—But First, Validate" by Daniel Solow of Case Western Reserve University (January–February). *www.amsci.org/node/5015*

Extended Interviews

The First Person interviews with Annie Kritcher (pages 211–212) and Cecelia Padilla-Iglesias (pages 235–236) only scratch the surface. To learn more about the role of models in their research, check out the companion podcasts on our website.

The Internet of Animals

Upgrades in global positioning system satellites, the use of the internet to link datasets and process information in real time, and the miniaturization of devices has allowed for more wild animal monitoring than ever before. Roland Kays of North Carolina State University discusses how this new Internet of Animals provides insights into animal migration and behavior, and also how data from animals can feed back into weather and environmental monitoring. A video of the talk and social media highlights are available on our website.



NASA/JPL/Caltech

Movebank/422 South

Look for this icon on articles with associated podcasts online.

American Scientist (ISSN 0003-0996) is published bimonthly by Sigma Xi, The Scientific Research Honor Society, P.O. Box 13975, Research Triangle Park, NC 27709 (919-549-0097). Newsstand single copy $5.95. Back issues $7.95 per copy for 1st class mailing. U.S. subscriptions: one year print or digital $30, print and digital $36. Canadian subscriptions: one year print $38, digital $30; other foreign subscriptions: one year print $46, digital $30. Print institutional rate: $75; Canadian $83; other foreign $91. Digital site license $200, print and digital institutional rate $275. Periodicals postage paid at Durham, NC, and additional mailing offices. Postmaster: Send change of address form 3579 to American Scientist, P.O. Box 193, Congers, NY 10920. Canadian publications mail agreement no. 40040263.

Kevin Heng | Mathematical models can address specific questions in scientific research, but their application requires intent and skill.

Approximating Reality

What does it mean to understand the natural world? To a classically trained physicist, it means that one is able to construct a model that not only accounts for currently measured phenomena but is also capable of predicting future phenomena. The modern challenge posed by artificial intelligence is that if one has a sufficiently large dataset of a system, in principle one may train a machine learning model to predict future phenomena without having any understanding of the underlying physical mechanisms. It reignites the debate of what "understanding" really means.

In a mechanistic view of the world, any phenomenon may be understood as a large ensemble of interacting billiard balls. The reductionist asserts that as long as enough basic units ("billiard balls") are present in such a system, one may model or simulate it exhaustively. For example, if one wishes to understand how waves behave in the ocean, one only has to simulate the interactions of each and every constituent molecule of water.

This reductionistic approach has been extremely successful at describing nature and making useful predictions, but it does not satisfy a more comprehensive meaning of "understanding." Even if one could run such a simulation in one's lifetime, it would be difficult to identify the underlying mechanisms responsible for producing different types of waves. Generally, examining the microscopic components of a complex system in order to understand how macroscopic behavior emerges falls short—think biology and economics. Implicitly, this brute-force approach suggests that one may study complex systems without having a scientific question in mind, and if enough computing power is deployed then insight naturally emerges. The billion-euro Human Brain Project is a spectacular example of the limitations of such an approach. Essentially, their simulations have failed to replace laboratory experiments.

As has been noted by the climate scientist Isaac Held, there is a tension between simulation and understanding. We simulate in order to mimic as much of an observed phenomenon as possible. But we achieve understanding by using idealized models to capture the essence of the phenomenon. By design, idealized models necessarily employ assumptions.

Models and Their Limitations

Models are approximations of reality, constructed by scientists to address specific questions of natural phenomena using the language of mathematics. Mathematical equations allow the practitioner to decide the level of abstraction needed to tackle a given problem, thus transcending the "billiard ball" approach. One example concerns the modeling of disease transmission. Human beings are complex entities whose individual behavior cannot be easily modeled by mathematical equations. However, the movement of ensembles of human beings within a city or country may be approximated by a set of fluid equations: the so-called compartmental models of disease transmission. Our incomplete understanding of the biological properties of a pathogen, as well as how it mutates and is transmitted across hosts, is encoded within a single parameter known as the reproduction number. In this manner, epidemiologists were able to study COVID-19 during the pandemic without having full knowledge of it, because the focus was on its macroscopic behavior across large spaces. (*See "COVID-19 Models Demand an Abundance of Caution," April 23, 2020.*) In such an example, the theorist is employing the principle of separation of scale in order to isolate phenomena.

Ideally, one would like to construct a universal model that is able to answer every question that the scientist asks of it. In practice, mathematical solutions to equations describing nonlinear systems that can be written down on paper are exceedingly rare. (*Nonlinear* here means that changes between the inputs and outputs of the system do not obey a simple relationship.) Instead, one must solve these equations with a computer, which allows one to study how different components of a physical system interact and produce nonlinear out-

comes. A universal simulation would be an *emulation*—a perfect replica of the actual system across all scales and at all times.

In practice, simulations are often limited by a *dynamic range problem*, meaning that one runs out of computing power to simulate behavior at all scales. The discretization of mathematical equations in order to program them into a computer introduces subtleties such as *numerical dissipation*—an artificial side effect of differential equation calculations that does not arise from physics. It is also likely that one needs different numerical methods to simulate behavior on different scales, and a single method is insufficient for all spatial scales. Perhaps the advent of AI will allow us to overcome some of these limitations, but for now emulations remain an aspiration.

If an astrophysicist wishes to understand how stars evolve over cosmic timescales (billions of years), then it is not unreasonable to model stars as spherical objects. Over such long timescales, the practitioner is less interested in transient phenomena and more interested in time-averaged effects. However, if an astrophysicist is interested in how sound waves propagate across a star so that they might understand its detailed structure (and therefore ultimately infer the age of the star, known as *asteroseismology*), then spherical symmetry becomes a questionable approximation. In these examples, the design of the model is not only tied to the specific scientific question being asked, but also the characteristic timescale being considered. Essentially, models often operate under idealized conditions in order to produce useful answers.

Once one accepts that models and simulations are necessarily limited in scope and may pragmatically be built only to address specific questions, then one has to accept that approximations are a feature and not a bug. Some practitioners even regard them as an art form. The intent and skill level of the modeler becomes relevant, because one needs to be able to ask a sharply defined scientific question and construct a model so that its output may be decisively confronted by data. Issues such as falsification and Occam's razor become relevant—one wishes to construct models that are complex enough to include the relevant physics, chemistry, biology, and so on, but simple enough to be proven right or wrong by the best data available.

Understanding Mechanisms

Models and simulations have been used not only for prediction but also to seek insight into underlying mechanisms. Some practitioners find that incorporating their prior beliefs about a system as well as sources of uncertainty associated with the data—an approach called *Bayesian statistics*—facilitates this confrontation between models and data. In this process, the inferred values of the parameters of the model are probabilistic. Bayesian frameworks allow for empirical sources of uncertainty and partial theoretical ignorance, as well as *degeneracies* (different combinations of parameters that produce the same observable outcome), to be considered when using models to interpret data.

Once AI technology is capable of designing models, writing programs to compute them, and interpreting their outcomes, it will forever change the way researchers use computers.

Even if it were possible to construct a true emulation, though, strictly speaking one would produce only correlations between phenomena. As scientists, we are interested in cause and effect. We want to understand why phenomena change in a particular way, which leads to better predictions, decision-making, and overall conceptualization about how a process works. To transform correlations into causal relationships, one needs to construct a simplified model to explain trends observed in the emulation. Such a model necessarily involves judiciously taken approximations in order to isolate the causal mechanism responsible for the observed trend. It is almost as if we are missing part of the language of models and simulations, a tool of intermediate complexity that goes beyond written linear models and full-blown numerical simulations.

Perhaps advances in AI will allow us to overcome this limitation of the human imagination. Perhaps scientists must ultimately let go of the ideal of explaining phenomena from first principles and simply accept that complex systems are not always amenable to the classical, reductionistic approach. It is plausible that an AI procedure, akin to an advanced version of the scientific computing software Mathematica, could produce approximate mathematical solutions of a governing equation much faster than any human being could. It is conceivable that this AI could visualize correlations and causations in multiple dimensions, beyond what a human brain could. Once AI technology is capable of taking over the tasks of designing models, writing programs to compute them, and interpreting their outcomes, it will forever change the way that researchers use computers to solve complex problems of the physical world, as well as how they train future generations of graduate students.

For example, a senior researcher may no longer need to be an expert computer programmer and may instead work with an AI that essentially acts as a super research assistant. Graduate students may spend less time on software engineering and more time asking deep, creative questions of the models or simulations they are studying. In other words, our computers would be only a small step away from telling us not only how to solve a problem, but also which problems to solve and what their solutions imply. AI would then acquire two human traits that are thought to be elusive to computers: physical intuition (or simply having a "feel" for how something works) and insight. Ironically, AI may end up emphasizing the most precious human trait we have as researchers, which is to ask deep, insightful questions led only by our curiosity.

Kevin Heng is professor of astrophysics at the Ludwig Maximilian University in Munich, Germany, where he leads the Chair (sub-department) of Theoretical Astrophysics of Extrasolar Planets. He is honorary professor at the University of Warwick in the United Kingdom. He is author of the 2017 textbook, Exoplanetary Atmospheres: Theoretical Concepts and Foundations *(Princeton University Press). Email: Kevin.Heng@physik.lmu.de*

Wikimedia Commons

Anne Goujon | Demographers study the past and present to prepare us for the future.

Retirees sit and stroll along the Eastbourne Seafront in the United Kingdom. Demographers help policymakers anticipate future societal challenges, such as the rising health care needs of rapidly aging populations in countries across Europe.

Roger Cracknell 01/classic/Alamy Stock Photo

Predicting an Aging and Changing World

We live at a unique time of transition in human history. Populations in some parts of the world are growing rapidly, while others are stabilizing or declining. At the same time, age distributions are shifting at different rates in different regions. The populations of Niger and the Democratic Republic of the Congo, for example, are still quite young, whereas Italy and Japan are much older. Even within countries, there are growing gaps in population growth rates and age structures across urban, suburban, and rural areas.

These complexities confront us with daunting questions: How can we prepare countries for the huge social, cultural, political, and economic challenges that come with these major, often unprecedented changes? Increasingly, policymakers turn to the field of demography for answers.

As a demographer, my job is to look to the past and present to understand how variables such as fertility, health, and education cause shifts in the size and age distributions of populations—and, ultimately, how they affect the future. To a large extent, demography is destiny: The trends of the next few decades are in motion right now. We cannot change what is already done, but by understanding the factors that shape long-term population changes, demographers can help policymakers steer the trajectories of the future, while preparing for inevitable challenges ahead.

An Aging Europe

Two of the biggest issues facing many European countries are populations that are both decreasing in numbers and aging overall. Policymakers need to find ways to provide care for the growing number of older people while also sustaining an economic system based on a smaller population of working age individuals. In Italy, for example, greater than 30 percent of the population is over age 60, while the current age at retirement is 62. That trend will place an intense strain on government services and the tax revenue used to support them. And Italy's case is far from unique.

It turns out that all countries across the globe are on a similar trajectory, as they all go through what demographers call the *demographic transition*. The demographic transition is a theory, primarily developed by American demographer Frank Notestein in the 1940s, that describes changing patterns of fertility, mortality, and population growth as countries develop. Greatly simplified, traditional societies with high fertility and mortality rates expe-

rience little or no population growth; societies transitioning to modernity see a temporary but rapid increase in population as death rates drop before birth rates; and modern societies return to little or no population growth, due to low fertility and mortality rates.

Although all nations are on this same path, the demographic transition does not occur at the same time and pace everywhere. In more developed countries, the demographic transition typically happened over a long period, starting in the late 18th and early 19th centuries in many European countries and in the United States. In these countries, the demographic transition is now complete: The population is growing slowly, declining, or projected to begin declining within the next few years and on through the end of the century. (*See figure below.*)

As population numbers decline, the proportion of older people in a society increases. In 2020, while working at the European Commission's Joint Research Centre in Ispra, Italy, my colleagues and I studied aging patterns in Europe to help countries anticipate future challenges, such as health care and workforce needs. We found considerable variation across countries, with southern Europe aging faster than the rest of Europe. But all regions are expected to converge on this pattern, and aging will happen in both rural and urban areas. In regions experiencing rapid aging and depopulation, the loss of population can lead to declining infrastructure, reduced access to public services such as education and health care, and social disruption.

One factor that might buffer a society from this fate is migration, both from other countries and from other regions within a country, which brings in additional (and typically younger) people. Our analysis shows that most regions within the European Union have more people arriving than leaving. But inward mobility is only high enough to offset the decline in workers caused by people retiring in 27 percent of regions, and primarily in urban areas. In only a few areas is migration making up for the lack of young people entering the labor force.

To a large extent, demography is destiny: The trends of the next few decades are in motion right now.

The variation in demography across regions, even within individual nations, suggests that European and national leaders should tailor their policies to specific needs at the local level.

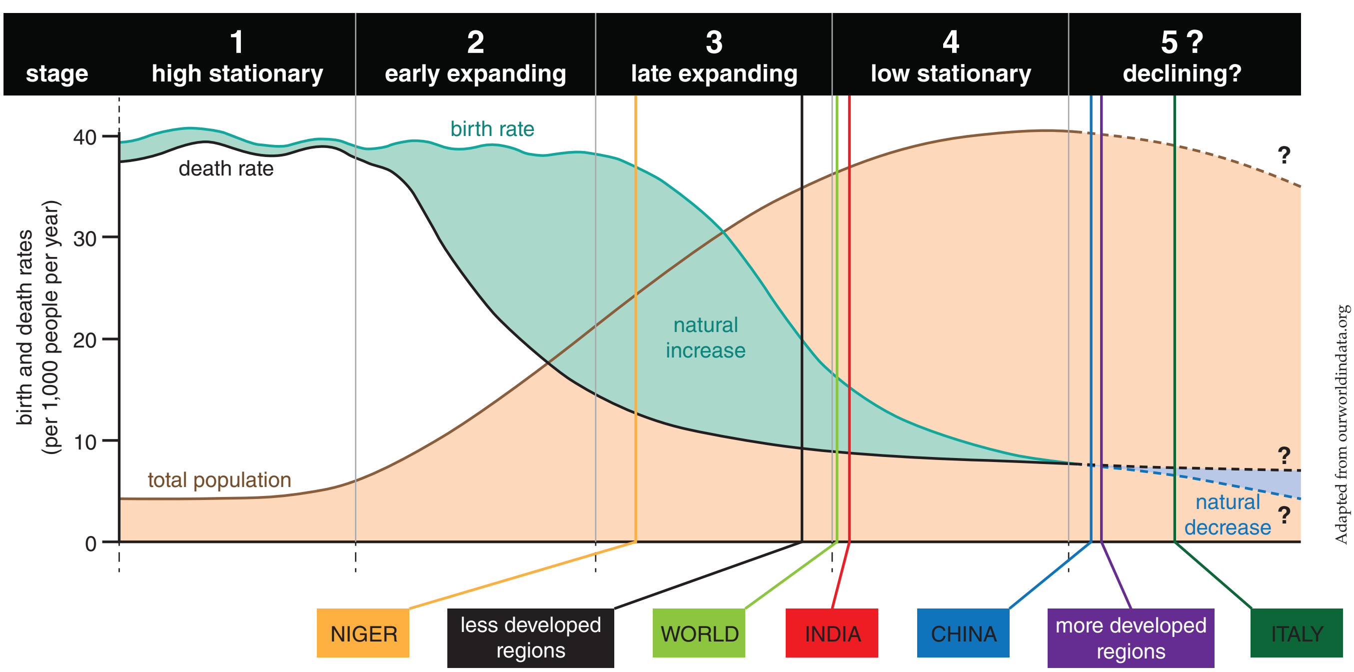
Education in Niger

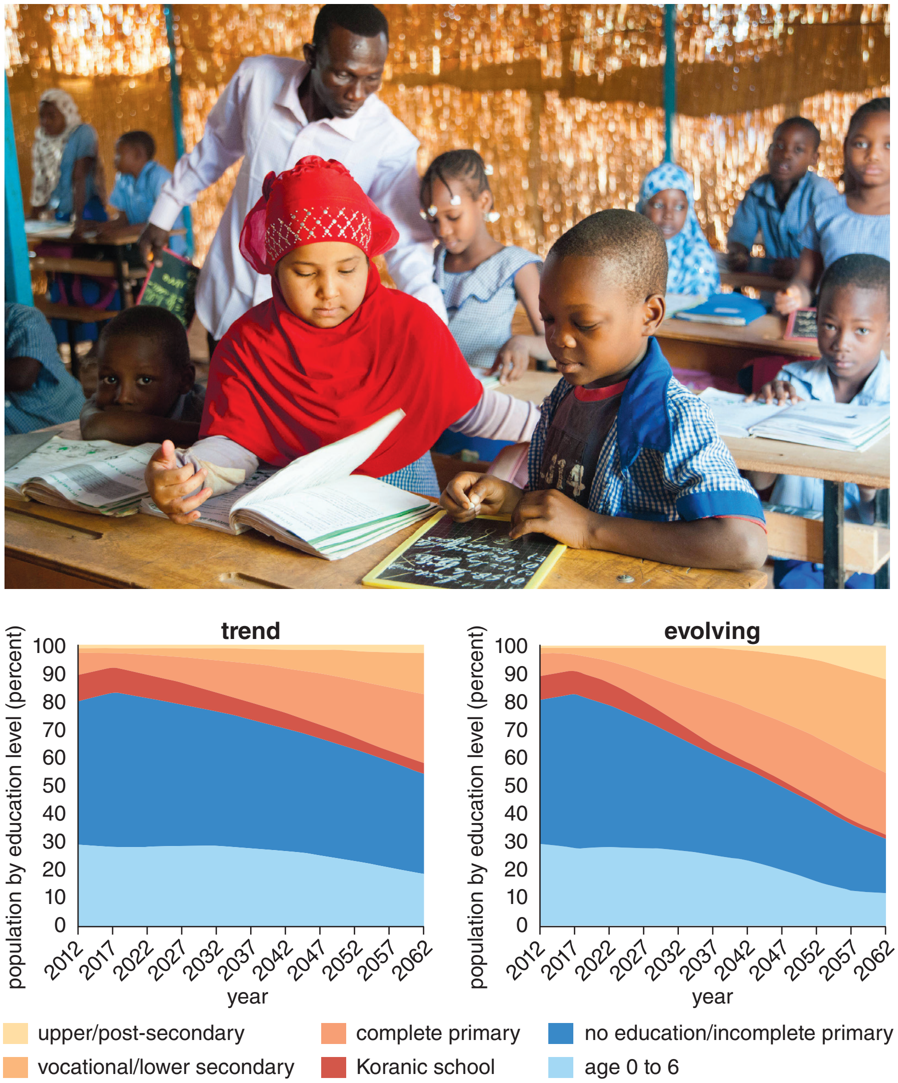
Meanwhile, less developed nations face a different near-term challenge: rapid population growth. The Democratic Republic of the Congo, for example, is growing at a yearly rate above 3 percent, with a population doubling time barely over 20 years. Policymakers in these countries must find ways to create enough health care, education, and employment opportunities to keep pace with growing numbers—all significant challenges in regions with relatively limited resources.

Fertility rates in developing countries are declining due to economic development, the spread of women's education, urbanization, increased access to contraception, and changes in social norms around marriage, family, and gender roles. Yet these countries are still in a phase of population growth because they began their demographic transitions much later than did more developed countries; for many parts of Africa, the demographic transition began only in the last quarter of the 20th century.

These transitions are now moving quickly, mostly because of medical and hygienic progress that took a long time to develop and spread. Even so, populations in less developed countries should keep increasing until the end of the 21st

The demographic transition theory predicts that all countries shift from rapid population growth to stable or even declining population size as death rates and then birth rates decline. Today, countries in later stages of development, such as China and Italy, are further along on this trajectory. Countries in earlier stages of development, such as Niger, have only recently begun to see birth rates begin to decline.





Adapted from Goujon et al., 2021; Alice Mutasa/Alamy Stock Photo

The author's projections (*bottom graphs*) suggest that improving funding for schools in Niger (*top*) could dramatically increase the proportion of educated citizens by 2062 (evolving scenario), whereas maintaining the status quo would result in more modest educational gains (trend scenario).

century, placing these countries, on average, 80 years apart from more developed regions in the demographic transition process. The last country that until the early 2010s had not started its transition to decreasing fertility was Niger in West Africa, which until recently had had an average fertility rate of greater than seven children per woman for several decades. But a survey published in 2021 shows a rapid decline to 6.2 children per woman.

In 2017, I worked with the United Nations Children's Fund (UNICEF) country office in Niger and with the Nigerian government to explore future demographic prospects and their development implications. Niger has the fastest population growth in the world and is also among the least-developed countries. A large majority of the population does not have any formal education. Niger's rapidly increasing population presents several issues, including economic pressure, food insecurity, and malnutrition. Most rural regions are particularly prone to drought, desertification, and environmental degradation, while also having limited access to education and health care.

In my role as a demographer at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, my team and I worked with local experts and stakeholders to develop narratives about the possible future of Niger, with an emphasis on the potential impact of educational development. We know from past demographic trends in other parts of the world that giving girls access to education is an effective way to curb population growth. Education also greatly increases a nation's economic prospects.

By considering different levels of investment in education, we could project different population growth and economic scenarios for Niger. We found that substantial population growth is inevitable within the next two decades, during which the population will likely double from 25 million in 2022 to 50 million to 58 million in 2042, depending on the scenario. The trends of the next few decades are already in motion, too advanced to be changed without drastic measures or unforeseen events. But this same inertia calls for immediate action to secure the future. We recommend a focus on curbing population growth, possibly through awareness campaigns and increased availability of family planning. We also recommend a greater investment in education, which could lead to lower fertility and increase the productivity of men and women in the country.

Assuming a continuation of present policies, by 2062, only 35 percent of 25- to 39-year-olds in Niger will have a secondary education or higher. But by increasing investment in education today, and boosting enrollment in vocational, primary, and secondary education, the proportion of citizens in this age group with a secondary education or higher would increase to above 70 percent by 2062. In this alternative scenario, Niger's population size by 2062 would triple to 78 million, rather than quadrupling to 106 million as expected under current policies. (*See figures at left.*)

The Future of Ukraine

Demography moves at a slow pace—except when shocks happen, such as a war. Such effects occurred on a huge scale during World War II. My colleagues and I see them happening in a different context now in Ukraine. Before the 2022 Russian invasion, the population of Ukraine was already shrinking due to relatively low fertility, high mortality, and especially high emigration rates. Now, the war has put Ukraine on a path of rapid population decline. Greater than eight million have fled—a loss of almost 20 percent of the total population.

Predicting Ukraine's future population size can help the region and the global community prepare for the country's postwar needs. At IIASA, my colleagues and I worked with other demographers at the European Commission's Joint Research Centre to develop four hypothetical scenarios for

Ukraine's demographic future, based on different assumptions about the duration of the war, the scale of displacement, and longer-term migration patterns, including postwar returns. Because the war is ongoing, these projections are very uncertain in absolute terms. But we can say that the impact of forced displacement and longer-term migration patterns on the future population size of Ukraine will be drastic.

In all four scenarios, Ukraine's population is projected to continue to decline until 2052. In the most pessimistic scenario, which predicts a long war and low numbers of Ukrainian nationals returning after the war, our models predict a 31 percent decline in population by 2052 from the start of the war. Even in an optimistic scenario where Ukraine recovers quickly from the war and becomes a country where in-migration outnumbers out-migration, the results suggest a population decline of 21 percent.

Our work indicates that the war and resulting displacement will exacerbate the decline and aging of the Ukrainian population, with dramatic changes in the population structure, particularly in the younger age groups. These projections indicate that policymakers will need to address the rapid aging of the population and the loss of human capital with strong strategies to educate the working-age population to improve lifelong skills, facilitate the reintegration of returning migrants and their families into society and the labor market, and engage with Ukrainian communities abroad to encourage their return, even for generations born in other countries.

An Uncertain Future

Demographic projections can also help us think about our human future at a planetary level. The world's population as a whole is aging and is on the verge of beginning to contract. This transition will mark the completion of the demographic transition at the global scale. But we do not know exactly when the global population will peak or how high it will reach. At the Wittgenstein Centre, where we at IIASA collaborate with demographers at the Austrian Academy of Sciences and the University of Vienna, there is a long tradition of developing global population projections. These models account for the impact of educational attainment. They also rest on assumptions about future fertility, mortality, and migration, which we develop through discussions with scientific experts. According to our forthcoming 2023 projections, the world population will reach 9 billion by 2040, peak in 2080 at 10.1 billion, and decline thereafter to reach 9.9 billion in 2100.

Other demographic institutions have arrived at somewhat different projections based on differences in their own underlying assumptions. For instance, the United Nations Population Division forecasts a slightly later peak and 10.4 billion people in the world in 2100. They predict that greater number largely because they assume slightly less rapid fertility decline in those countries that have, at the moment, high fertility. This difference in projections illustrates the many uncertainties built into any demographic model.

Demography moves at a slow pace—except when shocks happen, such as a war.

One major uncertainty is how low fertility can go. The economic, social, and policy factors that lead to low fertility levels interact in complex ways, and the reasons for low fertility rates may vary from country to country, making global predictions challenging. At the moment, fertility is extremely low in a few East Asia countries. South Korea and Hong Kong now average 0.8 children per woman. In 2016, China ended its one-child policy, but that change did not lead to the expected increase in fertility. The country currently averages only 1.1 children per woman. As a result, India surpassed China as the world's most populous nation earlier this year, and China's population could drop by roughly 50 percent by the end of the century.

A second uncertainty is how high life expectancy can go. Longevity depends on a variety of hard-to-predict factors, including advances in medical technology, changes in lifestyle and behavior, and improvements in public health measures. We know that life expectancy has increased significantly in many parts of the world over the past century. In Japan, for example, life expectancy at birth has increased from around 38 years in 1900 to around 85 years in 2022. Similar trends have been observed in many other countries. Some experts believe that life expectancy may continue to increase, possibly reaching as high as 100 years or more in some parts of the world by the end of the 21st century. Such extreme longevity would have profound demographic implications. But this prediction is quite speculative.

Climate change may influence demographic trends, but its impact is also unclear. Of particular concern is how a warming planet may drive mass migrations away from the most affected places, such as coastal areas, or trap poor populations living in conflict within hard-hit agricultural regions. The increased prevalence of sudden- and slow-onset disasters could also affect health, mortality, and fertility.

Despite these many uncertainties, demography gives us windows into a range of possible futures so we can prepare for whatever lies ahead. My work on education shows that total population numbers are not the only meaningful consideration. The capacity of humans to come up with innovations, solutions, and adaptive measures is equally important in addressing the stark contrasts between the availability of natural resources and the billions of humans who require them to sustain life.

Bibliography

Goujon, A., C. Jacobs-Crisioni, F. Natale, and C. Lavalle, eds. 2021. *The Demographic Landscape of EU Territories: Challenges and Opportunities in Diversely Ageing Regions*. Luxembourg: Publications Office of the European Union, JRC123046.

Goujon, A., G. Marois, and P. Sabourin. 2021. Deriving Niger's demographic and education future to 2062 with stakeholders: Which results? *Population Research and Policy Review* 40:617–627.

Ueffing, P., S. Adhikari, K. C. Samir, O. Poznyak, A. Goujon, and F. Natale. 2023. *Ukraine's Population Future after the Russian Invasion: The Role of Migration for Demographic Change*. Publications Office of the European Union. JRC132458.

United Nations, Department of Economic and Social Affairs, Population Division. 2022. *World Population Prospects 2022: Summary of Results*. www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf

Anne Goujon leads the Population and Just Societies (POPJUS) program at the International Institute for Applied Systems Analysis in Austria. POPJUS focuses on identifying sustainable development challenges and exploring people-centric systems solutions for sustainable, resilient, just, and equitable societies. Developing population projections are an important part of her work. Email: goujon@iiasa.ac.at

Bias Optimizers

AI tools such as ChatGPT appear to magnify some of humanity's worst qualities, and fixing those tendencies will be no easy task.

Damien Patrick Williams

Recently, I learned that men can sometimes be nurses and secretaries, but women can never be doctors or presidents. I also learned that Black people are more likely to owe money than to have it owed to them. And I learned that if you need disability assistance, you'll get more of it if you live in a facility than if you receive care at home.

At least, that is what I would believe if I accepted the sexist, racist, and misleading ableist pronouncements from today's new artificial intelligence systems. It has been less than a year since OpenAI released ChatGPT, and mere months since its GPT-4 update and Google's release of a competing AI chatbot, Bard. The creators of these systems promise they will make our lives easier, removing drudge work such as writing emails, filling out forms, and even writing code. But the bias programmed into these systems threatens to spread more prejudice into the world. AI-facilitated biases can affect who gets hired for what jobs, who gets believed as an expert in their field, and who is more likely to be targeted and prosecuted by police.

For some people, the word *bias* is synonymous with prejudice, a bigoted and closed-minded way of thinking that precludes new understanding. But bias also implies a set of fundamental values and expectations. For an AI system, bias may be a set of rules that allows a system or agent to achieve a biased goal.

Like all technologies, AI reflects human bias and values, but it also has an unusually great capacity to amplify them. This means we must be purposeful about how we build AI systems so that they amplify the values we want them to, rather than the ones accidentally fed into them. We have to ask questions about the source material that trains them, including books, social media posts, news and academic articles, and even police reports and patient information. We must also examine the frameworks into which that data is placed: What is the system doing with that data? Are some patterns or relationships between certain words or phrases given more value than others? Which ones? Why? What are the assumptions and values at play in the design of tools that transform human lived experiences into data, and that data into algorithms that impact human lives?

It is much easier to see through the mystique of ChatGPT and other AI applications once you understand exactly what they are and what they do. The truth about such algorithms is that they're literally just sets of instructions. You have a set of standardized operations within which particular weights and measures can be adjusted. In so doing, you have to adjust every element of the whole to make sure the final product still turns out the right way.

Algorithms are often sold as magical, but they are neither unexplainable nor even terribly unfamiliar. The recipe for any food—just as for anything you have to make—is an algorithm, too. My favorite algorithm is pumpkin pie. If you go to make a pumpkin pie, you might decide you'd like less butter, more sugar, or more milk. But you can't adjust the proportion of the pie's ingredients without considering the rest, or you'll end up with a crumbly or spongy mess; it won't really be a good pie. You must adjust the whole recipe, the whole algorithm.

To the person using it, an algorithm may look like a unitary thing that performs one job: A Google search, for instance, seems like a singular, powerful operation that searches the web. In reality, platforms and search engines work on dozens of algorithms that search, sort, rank, weight, associate, suggest, amplify, and suppress words, concepts, and content. Those algorithms work in concert, but when you take a matrix of algorithms and automate it, it looks as if your computer system is autonomous and self-directed. So it is with the new AI chatbots: They seem to deliver on "true artificial intelligence," a seductive idea that goes back to the dawn of the computer age, but they are actually composed of series of algorithms even more complex than the systems that came before.

A History of Bias

Since the 1940s, when mathematicians and cryptographers such as Joan Clarke, Jane Hughes, Pamela Rose, the other 8,000 women of Bletchley Park, and Alan Turing used early computer technology to break complex codes and help win World War II, people have wondered about the possibility of intelligence in digital computers. In the 1950s, computer researchers began to ask, "Can machines think?" And in

QUICK TAKE

It is not surprising that systems trained on biased source material would result in biased outputs, regardless of whether or not the original biases were intentional.

Many groups are already integrating AI tools into their public-facing interfaces, billing them as assistants and interventions to help people do their jobs more efficiently.

Generative pretrained transformers (GPTs) such as Bard and ChatGPT cannot recontextualize or independently seek out new information that contradicts their built-in assumptions.

Dusan Stankovic/iStock Photo

Creators pour their emotions and subjective reactions into the algorithms that guide our lives, making these complex systems as idiosyncratic, volatile, and biased as their creators are.

the 1960s, a rift formed between two camps of AI researchers at Dartmouth. One group focused on computation and cybernetics, which are feedback loops that mimic biological processes. The other group worked to replicate human neural networks in electronic form. Neither camp considered machine bodies, emotions, or socialization, however. These researchers firmly believed that the key to AI was to divorce any messy social factors from the purity of rationality and intellect.

As part of this work, scientists developed language models (LMs), a method of determining the probability of words connecting to each other based on context cues such as their starting letter and the preceding word. One of the earliest examples was ELIZA, a program developed by computer scientist Joseph Weizenbaum at the Massachusetts Institute of Technology in 1964. At first, ELIZA was meant to parody open-ended psychotherapy; the program would do things such as rephrasing the typed inputs from users as parroted questions rather than replying with any new thoughts. Even knowing that they were talking to a computer, users repeatedly formed emotional bonds with ELIZA, often in as little as one or two short conversations. Weizenbaum was astounded at what he called the "powerful delusional thinking" such a brief engagement could produce.

ELIZA was one of the first mainstream LMs, but the work didn't end there. The dream of AI grew up alongside the dream of natural language processing. Researchers working on natural language processing sought to combine linguistics, computer science, artificial neural networking, and AI to find ways for computers to interpret, process, and communicate in human-like, conversational language. In the 2010s, the models Global Vectors for Word Representation (GloVe) and Word2Vec were two of the foremost examples of natural language processing systems. They work by statistically mapping the relationships between words, embedding layers of associative meaning between them.

Early LMs could represent the semantic connections between words such as "dog" and "dig" or "plane" and "flight." These early programs used so-called machine learning, a process of encoding various elements of English language as data, and then training the system to hit particular predictive targets and to reinforce associations

Jan Halaska/Alamy Stock Photo

Predictive policing is supposed to forecast crime and recognize suspects, but the algorithms that drive it can seed the system with racial prejudice and other toxic misinformation.

between the data points. Those associations were then mapped as mathematical representations of how strongly they're associated. In a sense, they were complex auto-complete programs: They predicted the ways words are likely to be strung together based on the ways language is typically organized in books, stories, articles, and so on.

But Word2Vec and GloVe had two major problems. First, their outputs often contained prejudicial bias. This bias occurred because the most readily available language sets on which they could be trained included things such as the more than 600,000 emails generated by 158 employees of the Enron Corporation in the years before the company collapsed. This particular dataset was full to the brim with human beings communicating in bigoted, immoral, or even just unconsciously biased ways about certain groups of other humans. Within what researchers now call the "Enron Corpus," you can find people trading and rating pictures of women; slurs against anyone of perceived Muslim background; and stereotypical "jokes" about the sexual proclivities of Black and Asian people. Tools using this material replicated and iterated the same prejudices, resulting in outcomes such as automated résumé sorters rejecting the applications of women and certain minorities at higher rates than white men.

The second problem was that Word2Vec and GloVe could not map associations across larger reams of text. The number of associations they could make actually decreased the larger the quantity of text got. These models group related words into compact, easily embedded representations; repeated word clusters translate into more strongly related associations. Thus, the larger the corpus, the more difficulty these older programs have mapping connections across the whole text, rather than the small, repeated clusters. Using more text as input requires different solutions—and thus the transformer framework was born.

Birth of the Transformer

The "GPT" in ChatGPT stands for "generative pretrained transformer." Its name describes a system of interoperable algorithms that weigh, arrange, and create associative distributions of text. They're built on large language models (LLMs), a subtype of LMs developed over the past five years or so, with datasets millions, billions, and now even trillions of words in size. LLMs are trained through deep learning—multiple layers of machine learning operating on and refining one another.

LLMs and the applications that use them, much like the forerunner language-model systems, are a form of automated word association in which words and phrases known as "language corpora" are turned into mathematical representations known as "tokens." The system is then trained on the tokens to predict the association between them. Well-trained natural language processing systems can interact with and guide a human through any number of tasks, from navigating a website to completing a complicated administrative application—or so the theory goes.

This approach often appears to work. You can use GPTs to generate a short story, summarize a book, or even just have a conversation. When someone types in a collection of words, the transformer measures those words against the tokens, and then generates a collection of words and phrases in a particular form, all with a high likelihood of fidelity to what the user sought. But these new systems retain the same prejudicial problems as Word2Vec, only now those problems multiply faster and more extensively than ever before.

Prejudicial bias not only informs the input and output of these systems, but the very structures on which they are built. If Google image recognition is trained on more examples of cats than Black people; or if the testing group for a digital camera's blink detection includes no people of Asian descent; or if the very basis of photographic technology doesn't see dark skin very well, how can you possibly be surprised at the biased results?

Because of those embedded biases, predictive policing systems tied to algorithmic facial recognition regularly misidentify Black subjects and recommend over-policing in Black communities. Algorithmic benefits distribution systems meant to serve disabled populations are dependent on outdated notions about standards of care for disability, both in the training data and in the weights and operations within the models themselves. AI applications in health care and health insurance routinely recommend lower standards of care to already vulnerable and marginalized individuals and groups. Rua Williams at Purdue University and independent AI researcher Janelle C. Shane have shown that GPT checkers have problems with original text written by neurodivergent individuals. Entering such text into automated plagiarism-checking software, which already disadvantages disabled and otherwise marginalized students, has a high likelihood of producing harmful false positives—something ad-

mitted by automated plagiarism company Turnitin in May 2023.

In general, systems trained on the "natural language" people use on the internet when they talk about marginalized groups is likely to cast those groups as lesser. Expressions of prejudicial values and biases are not restricted to explicit slurs and threats of physical violence; they can also emerge more subtly as webs of ideas and beliefs that may show up in all kinds of speech, actions, and systems.

Such prejudices are inherent in the data used to train AI systems. The factual and structural wrongness is then reinforced as the AI tools then issue outputs that are labeled "objective" or "just math." These systems behave the way that they do because they encode prejudicial and even outright bigoted beliefs about other humans during training and use. When it comes to systems such as ChatGPT, these problems will only increase as they get more powerful and seem more "natural." Their ability to associate, exacerbate, and iterate on perceived patterns—the foundation of how LLMs work—will continue to increase the bias within them.

Because machine learning reinforces these processes, these technologies become confirmation bias optimizers. The systems generate responses that seem like factual answers in fluid language, but that output is just matching what it has been trained to associate as the most correct-seeming collection of tokens. GPTs do not care when they get something wrong or perpetuate a harmful prejudice. They are designed only to give you an answer you're statistically more likely to accept.

That innocent-sounding goal contains immense potential for harm. Imagine an AI that discerns the ethnicity of a patient from a set of x-rays, and then integrates it with another AI that consistently misdiagnoses signs of renal illness in Black patients—or with one that recommends lower standards of care. Now add a chat integration intended to help patients understand their diagnoses and treatment options. Then feed all that back to human doctors as suggestions and recommendations for how they should interact with the human patient sitting in front of them.

AI models have been called as revolutionary as the internet itself. They've also been compared to precocious children. But at present, these children are the spawn of hegemonic corporations fundamentally motivated by maximizing profit. Should we really give them the authority to control what we consider real knowledge in the world?

Self-evaluating for bias is not something most humans do well. Learning how to design, build, and train an algorithmic system to do it automatically is by no means a small task.

Rethinking the System

If generative AI systems such as ChatGPT and Bard are meant merely to reflect the world as it has been, then they are extremely well-suited to that task. But if we want them to help us make decisions toward a better future, one in which we're clear about which values we want in our technologies and our cultures, then we need to rethink everything about them.

We know that we can mitigate AI's replication and iteration of prejudicial bias by intentionally altering their weights and associative tokens. In colloquial terms, doing so would tell the system to model the world in a different way. To do that—to engage in a process known as "bias bracketing"—these systems would have to be built on a framework that constantly checks, rechecks, and reevaluates the associations it has, and one that actively seeks out alternative associations.

Self-evaluating for bias, including implicit bias, is not something that most humans do well. Learning how to design, build, and train an algorithmic system to do it automatically is by no means a small task. Before that work can begin, the builders will also have to confront the fact that even after mitigation, some form of bias will always be present.

We also need to take a step back and reconsider the question, "What are these tools meant to do?" and understand that human values, beliefs, and assumptions will always influence our answers. Used differently, GPTs could help us recognize and interrogate the biases in our language and our social structures, then generate new ideas, riffing and remixing from what already exists.

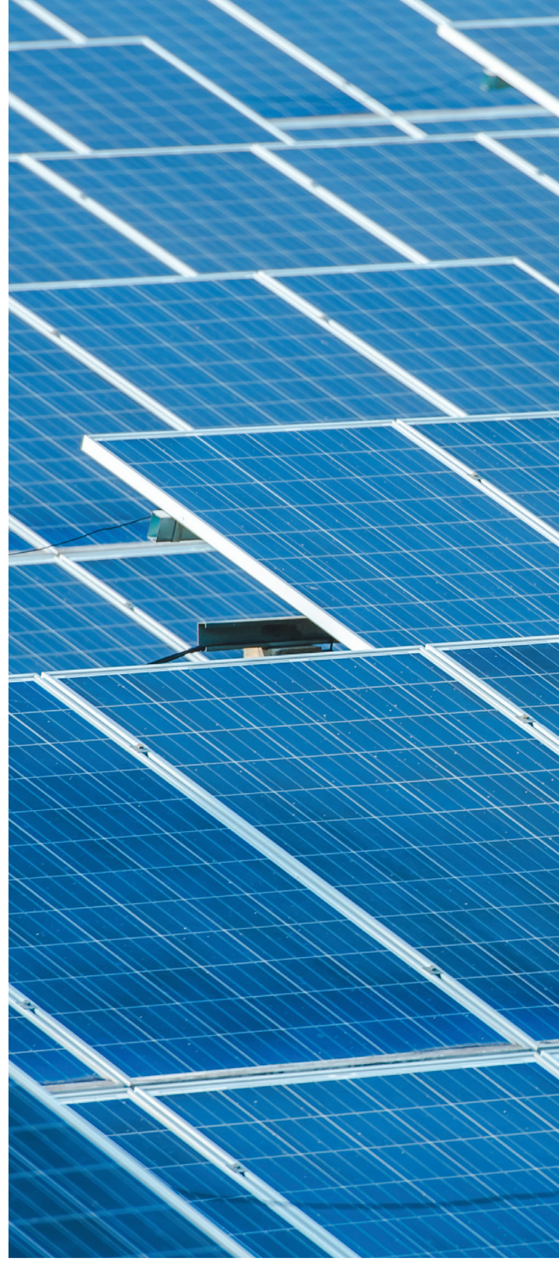
Imagine how much fairer and more constructive these tools might be if the data used to train them were sourced ethically from public domain works, or from people who volunteer their data, with a record of provenance, so we could be sure AI is not generating text or art that is essentially stolen from human creators. Imagine if GPTs had to obtain permission to use someone's data, and if data collection were opt-in rather than opt-out. Imagine how much more we could trust these systems if regulations required them to clearly state that they aren't truth-telling machines but are instead designed only to spit out collections of words that are statistically likely to jibe with our inputs. Imagine if the architectures of these GPT tools were shaped not primarily by corporate interests but by those most marginalized and most likely to be subject to and negatively impacted by them.

To build these systems differently will require more than a "pause" on development, as some AI researchers have repeatedly suggested. It will require AI system creators to be fully honest about what these systems are and what they do. It will require a reformulation of values, real oversight and regulation, and an ethic that sees marginalized people not as test subjects but as design leaders. Above all, it will require all of us to push hard against the prejudices that inform our creations.

Damien Patrick Williams is an assistant professor in philosophy and data science at the University of North Carolina at Charlotte, where he teaches and researches how technologies such as algorithms, machine intelligence, and biotechnologies are impacted by the values, knowledge systems, philosophical explorations, social structures, and even religious beliefs of human beings. Website: afutureworththinkingabout.com

Rupert Way | The economic case for an energy transition

Modeling a Greener Future



When scientists try to understand how Earth's climate is changing and how we might halt the rise of global temperatures, they turn to models. Some models predict how global temperatures will rise depending on future rates of greenhouse gas emissions, and how those temperature increases will likely affect humans and the environment. Other models forecast the costs and benefits of various strategies to mitigate climate change.

But all of these models have limitations and sometimes need to be adjusted based on new data. Now is one of those times.

The need to adjust models is especially acute when it comes to human and technological factors, which can be hard to quantify and about which we often have poor intuition. Over the past two decades, the half-dozen large energy models used to inform the influential reports by the Intergovernmental Panel on Climate Change have systematically overestimated future costs of key green energy technologies. For example, in 2010 a model used by the International Energy Agency projected that solar energy would cost $260 per megawatt-hour in 2020. The actual price that year was $50 per megawatt-hour, well below the average price for electricity generated by coal or gas. By overestimating green technology costs, these models have made the transition away from fossil fuel–based energy sources to green technologies appear substantially more expensive than it is likely to be.

Last year, my colleagues Doyne Farmer, Matthew Ives, and Penny Mealy, and I published a model that took a different approach to predicting the costs of four key green energy technologies: solar energy, wind energy, battery storage, and electrolyzers that combine electricity and water to produce hydrogen for use as a fuel or as a feedstock for other fuels. We reexamined the cost trends for these four green energy technologies and forecasted that they are likely to become much cheaper than most other energy models predict. We then made new projections for the expected cost of transitioning from fossil fuels to clean energy sources. Using our updated estimates, we calculated that transforming the global economy into a carbon neutral one by 2050 not only is economically achievable, but will likely produce trillions of dollars of net economic savings compared with continuing a fossil fuel–based system. If governments pursue smart policies, going green quickly is likely to be cheaper than either a slower approach or sticking with the status quo.

Over the past two decades, the large energy models used to inform influential reports have systematically overestimated future costs of key green energy technologies.

It should not come as a surprise that green energy technologies are on track to become cheaper in the future. Innovation and ingenuity have been driving down the costs of these technologies consistently for decades. The prices for solar energy, wind energy, and batteries have all dropped by more than 90 percent since they were commercialized in the 1980s and 1990s. Solar and wind are now the cheapest forms of new electricity generation to build, and it is cheaper to run an electric vehicle than one with an internal combustion engine. Why, then, do the major models keep overestimating the price of green energy? A group of assumptions embedded within the models seems to be the source of the problem.

Economists build models of the global energy system based on energy demand, resource availability, and past technological trends to predict how much different energy sources will cost in the coming decades. These large energy models typically include a few types of assumptions that constrain the values of variables at play. *Floor cost constraints* set hard limits below which costs for different technologies are not allowed to fall, to avoid

BELL KA PANG/Shutterstock

Current energy models overestimate the cost of green energy technologies such as solar power. A new model predicts that these costs will continue to drop, making the transition away from fossil fuel–based energy sources much cheaper than expected.

unrealistically low-cost projections. *Deployment rate constraints* set upper limits on how fast technologies can be manufactured and rolled out, to avoid predicting unrealistically rapid deployment. And *technology mix constraints* limit the amount of solar and wind power allowed in the electricity grid in the model, to reflect concerns about production gaps created when those green energy sources are not generating electricity, such as during the night and on days without wind.

In the standard energy models, all these constraints appear to have been set far too conservatively, systematically underestimating technological improvements. For example, in 2015 an influential model from one of the top international modeling teams limited solar and wind to providing a maximum of 20 percent of the power in the electricity grid. But in 2023, those sources routinely provide more than 60 percent of the power in large economies such as the United Kingdom and Germany. Similarly, before 2015 most major models included floor constraints for the costs to produce solar energy, the lowest of which was $750 per kilowatt. Today, the lowest-cost systems produce electricity for less than $600 per kilowatt. Setting pessimistic technology constraints has led to higher cost projections for renewable energy than what we actually observe.

Our model removes or greatly relaxes the assumptions built into the leading energy models, based on ideas that one of my colleagues, Doyne Farmer, a complexity scientist at the Institute for New Economic Thinking at the University of Oxford, developed in the late 2010s. Farmer started his work after talking to energy modelers to understand how they came up with technology cost forecasts that looked decades into the future. He was dissatisfied with what appeared to him to be highly unscientific methods, including their conservative assumptions about the future development of green technologies. Farmer thought that the forecasts seemed to rest on ideas that were not grounded in history or based on observations about how technologies change. Also, he noted that many of the ideas incorporated into the models had not been tested.

A Different Forecasting Approach

To develop more reliable, empirically grounded models, Farmer began by tackling the question of how well models can predict technological change. If he could answer that question, he could then in turn develop cost models that would more accurately reflect the ways that technologies progress over time, avoiding unrealistic constraints that have distorted the existing models.

Farmer and his colleagues looked at various types of models for forecasting technological change over time and developed a method to evaluate their performance. The team collected historical data about 50 different technologies—including laser diodes, televisions, aluminum, and neoprene rubber—and used the models to predict the technologies' costs at various points in the past. This so-called *backtesting* allowed the researchers to assess the accuracy of each model by comparing its predictions with what actually happened for each technology. This approach also led the team to a method for forecasting costs probabilistically: That is, they could generate ranges of likely costs, called *confidence intervals*, rather than single cost values for the technologies. Producing ranges of predictions helps inform researchers about how wrong they should expect their forecasts to be, which is essential when dealing with systems containing high levels of uncertainty.

In the end, the model that performed the best at predicting technology costs was a type known as the *experience curve model*. First developed in 1936 in the study of airplane manufacturing, this model describes a widely observed pattern, in which the cost of a technology drops at a constant rate each time the cumulative production of that technology doubles. For example, between the 100th widget produced and the 200th, costs might drop 20 percent; then, with the 400th widget, costs will have dropped another 20 percent. The essential point is that the more we produce a given technology, the better we get at producing it and the cheaper it becomes.

Making use of this experience curve model, my colleagues and I conducted probabilistic cost forecasting for solar energy, wind energy, batteries, and electrolyzers—our four key green technologies. We predicted their costs under three different future scenarios. Two of the scenarios involved a transition from a fossil fuel–dominated energy system to one in which widespread renewable electricity and battery storage power most buildings, transportation, and light industry, while also producing clean fuels for hard-to-electrify sectors such as aviation, shipping, and the production of steel, cement, and fertilizer.

In the first of these scenarios, the transition to green energy happens quickly, by 2050. The second occurs more slowly, happening by 2070. The third scenario features no transition at all, with global fossil fuel use remaining high until at least 2070 and with no reduction in greenhouse gas emissions during that period. Once we predicted the cost of each technology in each scenario, we added up the cost to produce all of the major types of energy on Earth, including fossil fuels, to get the cost for the global energy system. We then calculated what is known as the *present discounted value* of each scenario. This calculation involves weighting future costs less heavily than present costs in the sum. It is a standard technique in economics used to reflect the fact that people generally prefer high-likelihood benefits received today rather than potentially uncertain benefits at some point in the future.

The fast-transition scenario was actually the cheapest, coming in at around $5 trillion to $12 trillion less than the cost of the no-transition scenario.

Unlike predictions made by the leading large models, we found that the fast-transition scenario was actually the cheapest, coming in at around $5 trillion to $12 trillion less than the cost of the no-transition scenario, depending on how much weight we gave to cost savings in the future. (The slow transition is cheaper than no transition at all, but not as cheap as the fast transition.) A rapid response such as the fast transition seems quite possible, since it involves simply continuing green-energy growth rates at their current levels for the next decade.

In this study, we focused only on the cost of the global energy system to give an apples-to-apples comparison between green and fossil fuel energy costs. We didn't account for the costs associated with the effects of climate change, because it is harder to conceptualize and quantify those costs. But continued use of fossil fuels is projected to cause increasing economic damage from floods, droughts, wildfires, and other extreme weather in the coming decades, greatly amplifying the benefits of a quick transition to green energy. Altogether, the fast scenario is likely to save tens to hundreds of trillions of dollars.

In the eight months since our paper was published, many other reports have also noted the trend of dropping costs for green energy technologies, and have recommended faster action on climate change. The International Energy Agency's forecast for renewable energy sources suggested that through 2027, the world appears to be on a pathway similar to that in our fast transition scenario. The policy environment for a rapid green-energy transition has also improved considerably, with policymakers across the world working to find ways to lower energy costs and reduce emissions simultaneously. For example, the Inflation Reduction Act in the United States has spurred trillions of dollars of new investment in green technologies, and the European Union and others are implementing similar policies to avoid falling behind.

Although there is a very long way to go before we can halt the rise in global temperatures, the rollout of green energy technologies is moving much faster, and at much lower cost, than most models predicted. With the right policies supported by updated and improved models, the path to a cheaper, greener energy system by 2050 now looks achievable.

Bibliography

Way, R., M. Ives, P. Mealy, and J. D. Farmer. 2022. Empirically grounded technology forecasts and the energy transition. *Joule* 6:2057–2082.

Rupert Way is an associate at the Institute for New Economic Thinking and an honorary research associate at the School of Geography and the Environment at the University of Oxford. Email: rupert.way@smithschool.ox.ac.uk

How to Spark a Fusion Fire

Courtesy of Blaise Douros, Annie Kritcher

Andrea "Annie" Kritcher has one of the highest-pressure jobs in physics. At Lawrence Livermore National Laboratory, she is the team lead for Integrated Modeling, and she is principal designer for fusion-energy experiments at the National Ignition Facility, or NIF. Those experiments bombard a small pellet with 192 precisely timed laser beams, causing it to implode and reach pressures of more than 100 billion Earth atmospheres. The process is called inertial confinement fusion *(ICF); its goal is to live up to the facility's name and achieve* ignition, *a long-sought state in which fusion reactions put out more energy than they take in. In 2021, after more than a decade of trials, NIF briefly came close to ignition. Finally, on December 5, 2022, Kritcher and her team achieved ignition, raising new hopes for fusion as a practical energy source. Kritcher spoke with* American Scientist *special issue editor Corey S. Powell about the historic achievement. This interview has been edited for length and clarity.*

From your perspective, what makes fusion such a difficult problem? Why is it so hard to model, and then why is it so hard to control the experiment?

The thing that makes it so difficult to control and model is that we need extreme conditions. We need extreme temperatures. In ICF [*inertial confinement fusion*] we also need extreme densities. We're reaching pressures that are more than two times the center of the Sun, and temperatures that are more than five times the center of the Sun, in our experiments. We're making the most extreme plasma state that you can make on Earth. As you can imagine, since that's not been done before, there's quite a bit we don't know about the materials science.

In these experiments, laser beams enter and hit the inside of a *hohlraum* [a hollow cylinder made of heavy atoms such as gold or depleted uranium] and create a very intense radiation bath. We have to be able to model that condition, and then we also have to model the plasma conditions of the implosion as it's imploding. Inside this intense radiation bath sits a spherical capsule, and in our experiments it's made out of diamond. And inside of that spherical capsule sits deuterium and tritium fuel [two hydrogen isotopes].

When we make this intense radiation bath, it heats the outside of the capsule that holds the fuel. That heat explodes the outside of the capsule, which, in a rocket-like effect, sends the remaining capsule and the fuel inward and squeezes it. We're taking something the size of a BB and squeezing it down to a size roughly half the diameter of a human hair. We're squeezing all of this material down to extremely small volumes, extremely high pressures, and reaching very high temperatures. But as the material implodes, it's not just the end state that's hard to model. It's the entire implosion that's hard to model as well. There are many states along the equation of the density-temperature-pressure relationship that we don't understand about the materials we're trying to implode. There's no experimental data.

You are dealing with temperatures, pressures, and conditions that go beyond known physics. How do you model something when you're going into such unfamiliar regions?

We have some low-pressure data and single points of parameter space that we use to benchmark our equation of state models. Then we have our theory, our transport models, radiation transport models, and hydrodynamics models. We benchmark these models as we go to integrated experimental data. We don't have a lot of data to support the basic physics models, or to test them, but we have generated a lot of integrated measurements, integrated datasets. You do an experiment where all this complicated physics is happening, and then you do an after-the-shot simulation to try to match all the observables from that experiment. It's like an integrated check on how the modeling is doing.

Integrated modeling is basically trying to model the entire system, from the laser hitting the hohlraum to ignition and what comes out of it. It's modeling how the lasers interact with the gold and depleted-uranium cylinder, how that implodes the capsule, and how that happens symmetrically. And then modeling the plasma physics conditions in the center of the dense plasma that we create, and the diagnostic signatures that come out of that as well.

What difficulties do you encounter when you're trying to deal with all of these different parts of the modeling?

There's quite a bit of simultaneous optimization that has to occur. When NIF first started [in 2009], it went for the highest yielding potential design that we had. The highest potential gain design. And that design had issues. It was more susceptible to instabilities. It just didn't work. There are a lot of situations where your model says one thing, but when you actually go to do the experiments you find that you're missing physics, or you can't calculate that physics because you don't have the resolution, or there's something you don't understand.

Over the last several years we've been working to rebalance, optimizing what's good for the implosion and what's good for the hohlraum together. They're usually not the same thing. We want to increase the size of the implosion, but when you do that, it becomes a more massive target. If you don't have more laser energy to blow off the material and implode that extra mass, then the implosion goes slower, and if you can't get it going fast enough, you can't squeeze the material fast enough.

We don't have any more laser energy to drive the implosion, so we have to make the hohlraum that surrounds it more efficient. For a given amount

of laser energy we put in, we have to get to higher temperatures. There's a lot of work around trying to do that symmetrically. To make the hohlraum more efficient, we had to make it smaller compared to the capsule. That way, the laser beams don't pass by the capsule; they go where you want them to go to get a nice, uniform radiation bath that surrounds the capsule. There's quite a bit of modeling that went into defining each laser beam to get a symmetric radiation drive during the entire laser pulse.

How long do you have to keep your lasers focused on the target during your fusion experiments?

It's about nine nanoseconds. There's also a really small, 2-micron fill tube—for reference, a human hair is 50 or 60 microns in diameter—that goes into the capsule, which holds the deuterium-tritium fuel. That little fill tube can send a jet of capsule material into the hot plasma and radiate the energy away. It's asymmetric, but the bigger issue is that the capsule material is *high-Z* [made of heavy, high-density elements]. If it shoots into the plasma material, it very quickly radiates the energy away from the hot plasma and cools it down. That has been one of the biggest challenges to model.

The goal is to reach just the right conditions in the hot plasma so that the fusion takes over and heats the plasma itself. Without what we call *self-heating*, we would never reach ignition conditions in these experiments. Out of the fusion reaction comes a neutron and helium. The neutrons escape, but the helium gets reabsorbed. That process carries a lot of energy. It self-heats the plasma, very rapidly increases the temperature, and ignites the fuel. We're trying to get the capsule to implode, ignite, and stay together long enough that we can burn up as much of the fuel as possible before it explodes.

NIF had a promising fusion result in 2021, but then the next few runs couldn't reproduce it. Why was it so hard to build on that success?

That plasma was designed to ignite under really good experimental field conditions. Unfortunately, all of the targets that we shot had quality issues. Dust particles, even particles the size of bacteria, can get into the capsule in fabrication or fall on the top of the capsule and ruin the experiment. We can see it with our diagnostics. In the x-ray pictures of the hot plasma, we see this little contaminant radiating all the energy away. Unfortunately, in two or three of the experiments, that killed the implosion. In another one, we had a big unintentional asymmetry, which we call an *odd mode asymmetry*. There are 192 laser beams, half firing from the bottom and half from the top. If the top half lasers fire just a percent different than the bottom half, it can squeeze the implosion and have it shoot out to one side.

We have been making design changes to try to be more robust with regard to these issues. In September 2022, we conducted the first test with extra laser energy and a thicker ablator. Since it was the first test, it was hard to get symmetry right, and the material was squashed like a pancake. But even though it was squashed like a pancake, it still produced nearly as much yield as the 2021 experiment, which had been perfectly spherical. That test was a good indication that even if we had this big perturbation, we could still get high yields. The idea is to be more robust to the presence of flakes on the capsule or to asymmetries. That's what we did with the last design changes.

"We're trying to get the capsule to implode, ignite, and stay together long enough that we can burn up as much of the fuel as possible before it explodes."

You're pushing the experimental envelope and pushing the modeling envelope simultaneously. Have you learned new things on how to model plasma physics from watching and having more data to work from?

We have various ways that the data feeds back into the modeling. A nice example of that is in the hohlraum plasma. We had kinetic effects in the plasma that were resulting in diffusing part of the density of the plasma in the hohlraum. Based on the focused experiments that were done, we knew we had to include that physics in the modeling to accurately represent what was going on in the hohlraum. Some things we've figured out and some things we haven't. But a lot of the process is trying to update our material response understanding in these extreme conditions, which is a physics model, and figuring out what physics we have to include that's not in the models already. We're modeling a big system and there's microphysics going on. A lot of times we have to use reduced-order models. It's figuring out how to benchmark those reduced-order models to accurately represent the experiment.

You made headlines when your fusion reaction produced more energy than it took in from the laser, though NIF as a whole still ran at a huge energy loss. What would it take to achieve system-wide breakeven?

NIF wasn't designed to be an efficient fusion energy–generating laser. It was just designed to make it work. The wall plug energy of NIF is 350 megajoules, and we are generating about 3 megajoules of fusion output. We got more fusion out than laser energy on target, but a working fusion plant would need to significantly increase the efficiency of the wall plug to the laser energy. And we need higher gain designs. A realistic prototype fusion power plant would be a gain of 30 megajoules, and a realistic operating plant with newer laser systems would be a gain of 100 megajoules. Right now, we're at a gain of 1.5 megajoules. We're a lot closer than we used to be. What we did was proof of principle, but it's still quite a long way to go for fusion energy.

What might that path look like? How could the tests at NIF lead to a commercial fusion power plant?

There's a lot of technology and design improvements that would need to happen. Even if we got really good at making uniform targets—which is one of the reasons why it takes us so long right now to do experiments—the design would have to change to do fusion 10 times a second [necessary to generate continuous power]. After this experiment, we checked the box on the physics requirements. Now it's really hard engineering stuff. It's not going to happen tomorrow, but it also took a long time from the first flight to having commercial flights. I'm hopeful.

A companion podcast is available online at americanscientist.org.

Of Mouse Models and Humans

In a nondescript laboratory, a university researcher using mouse models to find new drugs to treat major depressive disorder finds an interesting potential drug target. She is so sure this drug target will work that she decides to go independent and seek funding to start her own company. She is given start-up funds by a Big Pharma company and she speeds up development of the drug as quickly as she can to make it ready for clinical trials. The drug passes all its trials with flying colors except for one: the final clinical trial in humans.

In this all-too-common scenario, taxpayers and sponsors might well argue that the years of research and millions of dollars in funding were wasted. Indeed, according to the National Center for Biotechnology Innovation, it takes 15 years and an average of $900 million to bring viable new drugs to market.

Mouse models have long been used in biomedical research, but in recent years, papers published in the *Proceedings of the National Academy of Sciences of the U.S.A.* (*PNAS*) claiming that mouse models poorly mimic human responses have fanned the flames of controversy in the biomedical community. It all began in 2013, when a paper published in *PNAS* that ultimately was covered in the *New York Times* maintained that mouse gene expression responses to inflammatory diseases such as those brought on by burns, trauma, and endotoxemia (the buildup of dead bacteria, often from the gut, in the bloodstream) poorly mimicked human responses and concluded that research using mouse models was therefore a waste of time and money. A year later, another paper reinterpreted the dataset and presented the opposite conclusion. It was both intriguing and difficult to believe that the two papers drew contradictory conclusions after analyzing the same datasets from mouse and human models.

Though animal models have clearly helped in developing drugs that treat addiction, heart disease, and much more (*see box on page 215*), some still argue that too many preclinical targets have failed to be effective for human use and that vast amounts of resources have therefore been wasted. Much of this disagreement stems from unconscious bias on the part of researchers. In the aforementioned papers, the authors who concluded that animal models were ineffective were mostly physicians, whereas those who concluded that the animal models were efficacious were basic research scientists. These implicit biases have driven an unnecessary wedge between the clinical and basic research communities.

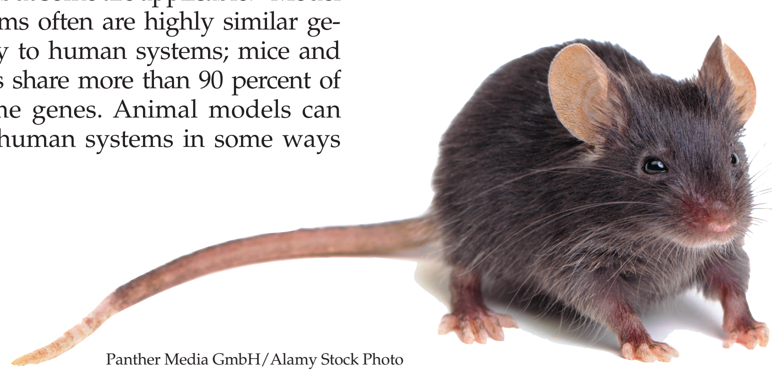
In a variation of George Box's assertion that "All models are wrong, but some are useful," we'd argue that "All biological models are imperfect, but some are applicable."

In a variation of statistician George Box's assertion that "All models are wrong, but some are useful," we'd argue that "All biological models are imperfect, but some are applicable." Model organisms often are highly similar genetically to human systems; mice and humans share more than 90 percent of the same genes. Animal models can mimic human systems in some ways while deviating from humans in others; the central question is the degree of congruence in the biological pathways of focus and whether the differences observed are associated with the research hypothesis. Any attempt to universally nullify or endorse animal model usage will benefit neither scientific discovery nor, ultimately, human health.

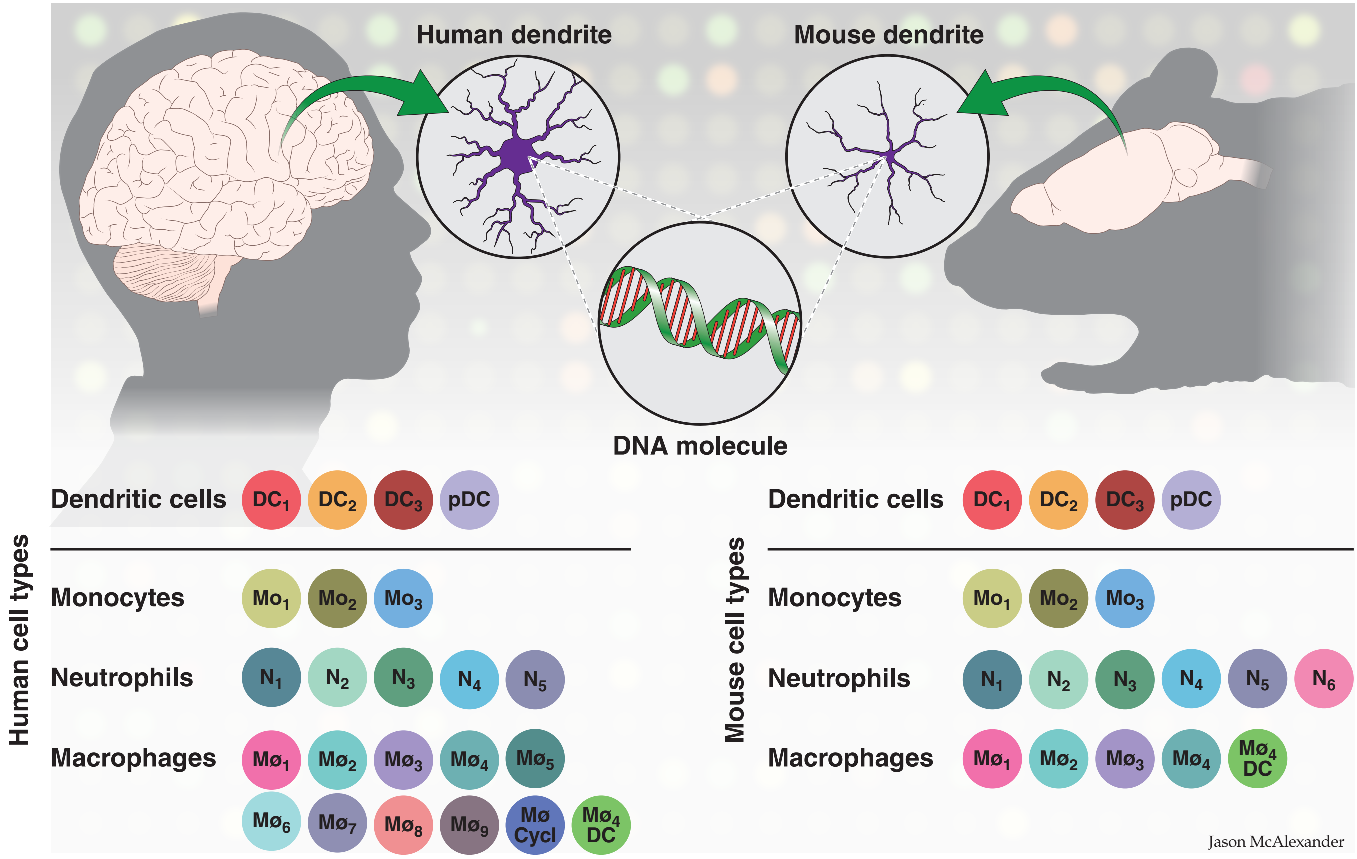
The World Needs Congruence

In many biomedical fields, it's common to see later studies contradict original findings and generate refutations. This phenomenon is known as the *Proteus phenomenon*, after the Greek god of change. However, any data scientist—statistician, computer scientist, or bioinformatician—cannot accept contradictory conclusions drawn from a single dataset; it threatens the rigor and credibility of the data science field.

Scientific research has more tools than ever at its disposal, but the deluge of information can be overwhelming and difficult to properly interpret. What we need to do first is to ask the right questions and then use the best tools available to interpret the data. To bridge the gaps between clinicians and basic researchers, we've developed a software program called Congruence Analysis of Model Organisms (CAMO). CAMO uses sophisticated statistical tools, aided by machine learning, to compare organisms at the molecular level and pinpoint disease-driving genes. This mechanism allows researchers to take a more granular ap-



Panther Media GmbH/Alamy Stock Photo



Mice and humans share more than 90 percent of their DNA and have many cell types in common. Researchers can use software tools such as Congruence Analysis of Model Organisms (CAMO) to find congruence between the two species, which can ultimately lead to more efficient discovery of potential drug targets.

proach, using congruence between animal models and human systems to find useful treatments for disease. For example, when we reanalyzed the human-mouse inflammatory disease data from the original papers that caused such controversy, we found mice were like humans in immune- and inflammation-related pathways and gene markers but different in the ways they translate proteins. This revelation could help researchers focus their attention where it's most needed and could also help rule out ineffective potential drugs earlier in the discovery process.

CAMO also provides enough statistical sensitivity to assess whether the data will even provide enough information to draw congruence conclusions in the first place. When sample size is small or biological variability with a species is large, CAMO warns researchers of insufficient information.

The Future Is Congruent

Returning to the researcher with start-up funding, we wish we could have provided her with CAMO early on in her research. Perhaps it would have helped her find better drug targets faster or helped her to realize that mouse models do not match humans closely enough in this instance to provide a good drug target. CAMO could likely have saved her many hours and millions of dollars and helped her bring more successful targets to trial.

Modern biomedical research offers a multitude of useful tools, but researchers also must learn to adjust their expectations and ask the right questions with those tools. CAMO is one tool that may help by showing where mouse models and human systems are alike enough to prove medically useful and where they are different enough that further research isn't indicated. Effective integration of different types of experimental data at all levels helps us grasp the bigger picture when we evaluate the usefulness of animal models. Ultimately, it's our hope that such tools will help usher in a new era of biomedical research, allowing us to bridge the gap between asking the right questions and finding the best answers.

Bibliography

Ahn, A. C., M. Tewari, C. S. Poon, and R. S. Phillips. 2006. The limits of reductionism in medicine: Could systems biology offer an alternative? *PLoS Medicine* 3(6):e208.

Bryda, E. C. 2013. The mighty mouse: The impact of rodents on advances in biomedical research. *Missouri Medicine* 110:207–211.

Ericsson, A. C., M. J. Crim, and C. L. Franklin. 2013. A brief history of animal modeling. *Missouri Medicine* 110:201–205.

Seok, J., et al. 2013. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences of the U.S.A.* 110:3507–3512.

Takao, K., and T. Miyakawa. 2015. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences of the U.S.A.* 112:1167–1172.

Zong, W., et al. 2023. Transcriptomic congruence analysis for evaluating model organisms. *Proceedings of the National Academy of Sciences of the U.S.A.* 120:e2202584120.

George Tseng is a professor and vice chair for research in the department of biostatistics at the University of Pittsburgh School of Public Health. Tianzhou (Charles) Ma is an assistant professor of biostatistics in the department of epidemiology and biostatistics at the University of Maryland School of Public Health. Jian Zou received his PhD from the department of biostatistics at the University of Pittsburgh and will join the department of statistics and management in the School of Public Health at Chongqing Medical University in China. Email for Tseng: ctseng@pitt.edu

History and Ethics of Animal Model Usage

Humans have a long history of using animal models to learn about ourselves. The first records of animal model usage date back to ancient Greece. In the fifth century BCE, Alcmaeon of Croton observed connections between the brain and sensory organs in dogs. A couple of centuries later, Aristotle studied embryo growth in chicks.

Since then, animal models have played an indispensable role in biomedical research, leading to many of the biggest medical breakthroughs. William Harvey, founder of modern physiology, discovered blood circulation in the early 1600s after studying the anatomy of several species of animals. Surgeon Frederick Banting and his student Charles Best found that injections of pancreatic cell extracts relieved diabetic symptoms in dogs, leading to the discovery of insulin in the 1920s. The Salk and Sabin polio vaccine was developed based on more than 40 years of experiments using monkey, rat, and mouse models, leading to a successful double-blind trial on 1.8 million children in 1954.

Despite these obvious benefits, the use of animals in biomedical research has been a subject of debate and controversy for decades, as people have sought to balance medical advancements with animal rights. Today, most scientists accept that animals should be used for research ethically and responsibly. Research regulations and academic guidelines generally follow the principles of the Three Rs: Replacement of animals where possible, Refinement, and Reduction.

Many researchers have also begun questioning long-held assumptions about the validity and cost-effectiveness of using animal models. In 1929, Nobel Laureate August Krogh wrote in the *American Journal of Physiology* that "for such a large number of problems there will be some animal of choice . . . on which it can be most conveniently studied," a concept known as Krogh's principle. That principle no longer seems so certain, as many biomedical investigations have shifted from anatomy and physiology to complex models of understanding, managing, and treating disease.

Science Source

Galen was the physician to Emperor Marcus Aurelius of Rome in the second century CE. He studied anatomy and physiology by dissecting pigs, focusing particularly on the functions of the kidneys and spinal cord. Known as the father of modern medicine, he made many contributions in the fields of anatomy, physiology, pathology, pharmacology, and neurology.

In cancer, only 15 percent of drug candidates survive to Phase III trials in humans after successful preclinical animal studies.

Modern styles of medical research require a much higher level of congruence between animal models and the specific human diseases they are used to study. In cancer, only 15 percent of drug candidates survive to Phase III trials in humans after successful preclinical animal studies, and then only half of those are eventually approved for clinical use—a success rate of less than 8 percent. For psychiatric disorders such as depression, bipolar disorder, and schizophrenia, many drugs tested in animal models failed to produce results adequate for clinical use. Despite the advances of modern molecular biology, many of these diseases continue to stump clinicians due to the involvement of advanced neural networks and the potential for drug targets to produce toxicity. The pharmaceutical industry has therefore reduced its research and development of psychiatric drug targets over the past 50 years. Understanding the congruence of animal models to human systems may increase the success rate in drug discovery and recover the investments in drug development.

Building Better Growth Curves

Current standards for assessing growth in infants and children often raise unwarranted concerns. Better models could improve care.

William E. Bennett, Jr

I first met Michael years ago, when he was just over a year old, the first and only child of a young Amish couple. He was admitted to our children's hospital to figure out why he was growing so slowly. I was the supervising gastroenterologist that week, but he had already seen several of my colleagues, including other specialists. His diagnosis was something doctors call *failure to thrive*, a term plucked out of Victorian-era medicine that has no clear consensus definition and is simply what we say when a child isn't growing as we expect.

In Michael's case, this diagnosis was complicated by a rare genetic disorder caused by the deletion of a large portion of one of his chromosomes. Only a half dozen cases of this specific deletion have been reported, so little was known about the expected course of his life, other than that he would have substantial delays in development and might never learn to speak or walk. Most clinicians will never see this specific genetic problem once during their entire careers.

Despite this complexity—and perhaps because of it—Michael's medical team felt something must be done. A battery of tests was planned, including endoscopy and surgical placement of a gastrostomy tube, which would feed him more calories.

However, Michael's parents did not agree with the assessment of his doctors. They saw a small, healthy boy who didn't eat much but didn't seem to be thin or starving either. They saw a happy, vibrant child who was full of energy and was thriving. The family had many questions and were hesitant to proceed with any invasive tests or procedures. Michael's parents were correct, it turns out. After reviewing his records, I agreed with them: He was healthy and thriving.

So where did this dissonance between family and medical team arise? The answer lies in a ubiquitous tool of pediatrics, the *growth curve*. It's at the front of every chart in every pediatric clinic in the world. During every visit to a pediatrician or family doctor, starting soon after birth, a child's height and weight are plotted on a standard curve. Medical students and residents see hundreds of growth charts during their training, and reviewing them is an essential part of both pediatric checkups and the assessment of many medical problems in children.

But like most health information, context is supremely important when assessing a growth curve, and Michael's case illustrates the many weaknesses of this ubiquitous medical tool. Even for kids without genetic disorders, the problem is pervasive. I've spoken with hundreds of families during my career who have been told by their pediatrician or other provider that their child has failure to thrive—only to end up with a shrug of the shoulders from the same doctor after all the tests come back normal. The variation in growth trajectory, body habitus, and height seen in the healthy general population is enough on its own to throw growth curves for a loop.

Even though context is so important in assessing growth, we fail to teach trainees how these curves are made and why this process can be misleading. We also fail to understand the test characteristics—such as the false positive rate—in a tool we use every day. We fail to contextualize genetics and medical history sufficiently when making a binary decision, asking, "Is this a health problem?" And perhaps most heartbreaking, we often fail to look at the patient and talk to the family when arriving at a diagnosis. Unnecessary testing, expense, time, parental anxiety, and even risks for the patient are the usual results.

Although I have had a clinical and research interest in these problems for two decades, they never really hit home until I experienced them firsthand as a parent. Like Michael, my daughter has a rare chromosome

QUICK TAKE

Growth curves are a standard screening tool in pediatric clinics around the world. However, those using these curves end up flagging healthy kids as having potential problems.

Failing to understand the patient's context when they are not growing as expected can lead to unnecessary testing, expense, time, parental anxiety, and even risks to the patient.

Improving growth curves is no easy task, requiring the use of longitudinal data and more diverse datasets. Personalizing growth curves is theoretically possible with machine learning.

Liz Roll/Alamy Stock Photo

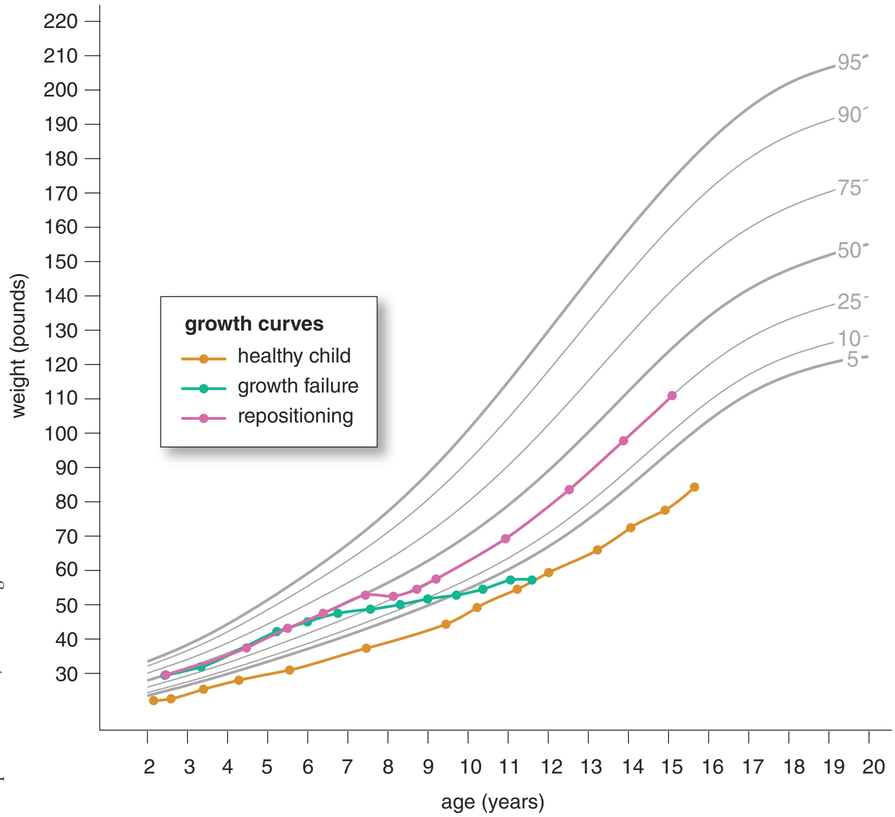
From birth onward, children are weighed at every health checkup, and their weight is plotted along with their height on standard growth curves published by the U.S. Centers for Disease Control and Prevention and the World Health Organization. But without understanding the limitations of these growth curves, they can result in a high false positive rate for diagnosing *failure to thrive*, a term indicating the child is not gaining weight or growing appropriately.

deletion—hers is called Phelan–McDermid syndrome. Like Michael, most children with her condition have profound delays in development, communication, and intellectual ability. And like Michael, her genes are telling her body to grow at a different pace than the growth curve suggests.

This problem is fixable. We can teach clinicians to use growth curves appropriately, and thus limit spurious diagnoses of failure to thrive and its attendant overuse of tests and procedures. We can use sophisticated tools in data science to enhance this mid-20th-century instrument. We can make the curves more inclusive and representative of human diversity, more accurate at predicting illness, and more personalized. We can build better growth curves.

Michael and Children like Him

Michael continues to thrive, years after I met him. The day I first spoke



Adapted from CDC; individual growth data from the author

Three real-life growth curves are shown plotted with the standard curves in gray, ranging from the 5th to the 95th percentile. The orange curve shows the growth of a healthy child whose rate of weight gain is far lower than normal. The green curve shows a child who is indeed undernourished and needs monitoring and health care guidance. The pink curve shows a healthy child whose growth follows one percentile curve for a while and then drops to another.

with his family, I reviewed the growth curve and what I had observed during my physical exam, and we elected to defer further workup. For the next year, we just watched his weight closely. He has never caught up to the curve and likely never will. But he has followed his own curve well below the first percentile, and he has stuck to it. Perhaps more accurately stated, his genes have stuck to it.

Conventional medical wisdom tends to accept many false positives in our quest to keep people alive.

A large change in genetic material, such as a chromosome deletion, invariably affects multiple body systems—and almost always the brain. One of the reasons that most children with major chromosome abnormalities have developmental delays is because of the sheer number of genes involved with brain development. As our primary evolutionary asset, our big brains require an enormous number of complex interactions, so neurodevelopment and neurological function utilize 80 percent to 95 percent of our genes. Missing genes or abnormal amounts of activation of these genes can affect development at any stage.

Growth has a similar, if not quite so dramatic, pattern. In a single-celled eukaryote such as the fungus *Saccharomyces*, 2,000 of the roughly 5,000 total genes are involved with growth rate, and about half of those are critical to growth. Of course, with complex multicellular organisms like humans, growth involves a large number of complex gene-gene interactions, and we are nowhere near understanding them all. One study in *Nature Genetics* found about 2 percent of the human genome is specifically geared to determining height alone, which is just one aspect of growth. The acquisition of body mass is likely more complex and involves much of the human genome's metabolic machinery, so much higher portions of our genes are involved, perhaps as much as 10 to 20 percent. One can see why a deletion of a sizable chunk of a chromosome will almost always include something critical to growth in children.

Based on these insights alone, we can deduce that some individuals can have variations that might not fit the typical growth curve.

How We Use Growth Curves

The first part of the problem with growth curves is how health providers use them, even in the best of circumstances. They are, in most cases, a screening test. That is, every child is monitored using growth curves, which are designed to detect growth problems with high sensitivity. But, just like many other screening tests, doctors frequently encounter false positives.

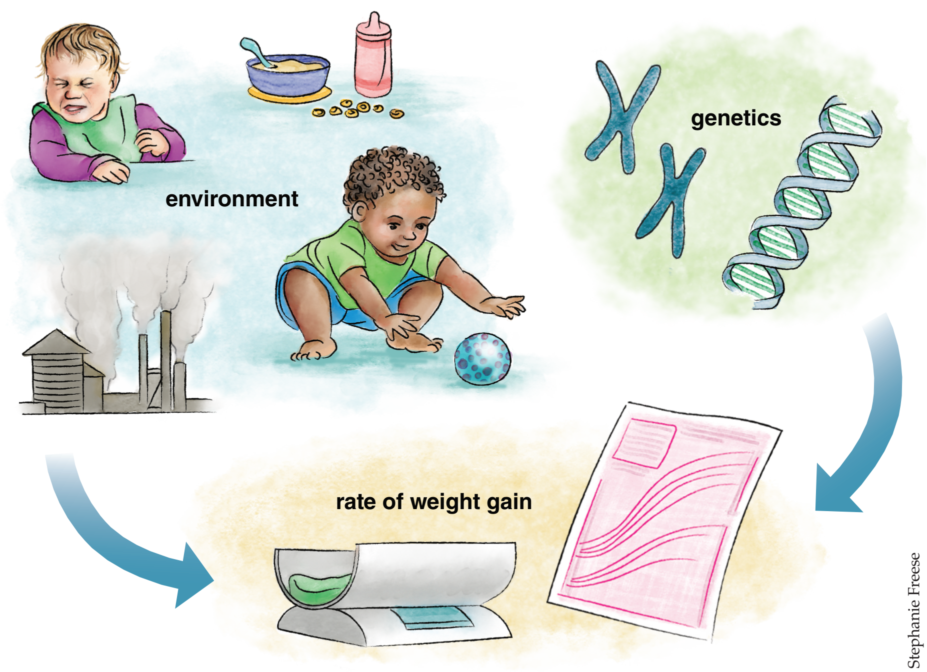
False positives are a pernicious dilemma in medicine. We want to make sure nothing slips through detection, because we are taught that it's much more serious to overlook a cancer or an infection (false negative) than it is to be wrong in retrospect (false positive). So where do we draw the line?

Conventional medical wisdom tends to separate tests into those for diagnosis and those for screening. For instance, the hemoglobin A1c blood test for diabetes is more than 99 percent specific if it finds elevated levels of glucose, but only about 30 percent sensitive. It therefore works great as a diagnostic test. But that's not the same as a screening test. Diagnostic tests assume you're targeting a population with symptoms. Screening assumes you're testing everyone, in a population with very low prevalence of the problem in question. Growth curves are screening tests. Most kids grow just fine, so they are more likely to have a measurement error or normal deviation from the curve than they are to have a serious underlying disorder causing growth failure.

Pediatricians use growth curves just like they do any other screening

test. When a child's curve starts deviating from the standard curve, that triggers the pediatrician to ask more questions: What is she eating? Does she vomit or have diarrhea? How is his development?

Specialists, such as gastroenterologists and endocrinologists, approach growth curves from a slightly different angle—for us, it's a diagnostic test rather than a screening test, because we don't do well-child visits. Most patients we see are referred by their pediatricians, and they've already had some period with a growth curve that has either fallen or flattened. Many have been told they have failure to thrive. Our job then is to assess whether they truly have growth failure, and if so, what has caused it. But even among that population, most patients I see referred for failure to thrive are completely healthy. In the meantime, though, many of these perfectly healthy children have undergone a variety of potentially unnecessary or harmful interventions—feeding tubes, excessive calories and force-feeding, radiographs that involve unnecessary radiation, potentially harmful surgical procedures, and disruptions in breastfeeding, among many others. Furthermore, the effects on family dynamics, parental anxiety, and parents' perception of their child's overall health can be drastic and hard to reverse.



Environmental context and genetics both influence a child's growth rate in complex ways. The acquisition of body mass involves much of the human genome's metabolic machinery, and the ways those genes interact with one another, the environment, and epigenetics are not well understood. Smaller parents tend to have smaller children, but diet, energy expenditure, pollution, stress, and many other environmental variables can play a role. Children with chromosome abnormalities also often have abnormal growth.

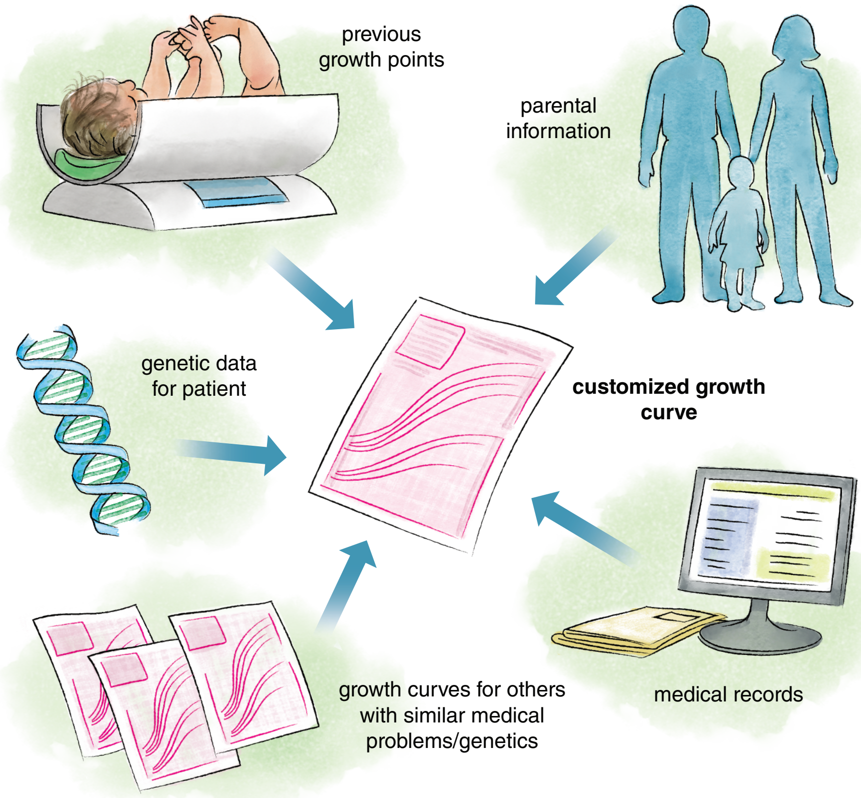
Why Curves Fail

Growth curves do not effectively identify problems, and they are not good at identifying who is "thriving" and doing well, either. These problems stem from the lack of a clear definition for the diagnosis for which growth curves screen: *failure to thrive*. The first mention of the term in the medical literature in 1906 is by Meinhard von Pfaundler, an Austrian pediatrician and director of the children's hospital in Munich. He used the term in an article about what was then called "maternal deprivation syndrome"—when neglected, abandoned, or orphaned infants showed growth rates and development far below other children. The term evolved over time, and in the 1930s it began to be used to indicate undernutrition or lack of adequate growth from any cause. Today, the term failure to thrive is a ubiquitous fixture of medical education. Various sources have since defined failure to thrive as falling below the third percentile or the fifth percentile, or crossing two or more percentile lines, or no growth for six months, or no growth for two months, or a growth rate below 20 grams per day, or any weight loss during childhood. You get the picture.

Another common problem is called *repositioning*. Many children will grow for months or years along one line on the growth curve, then inexplicably move to another percentile line and grow along it instead. This shift happens commonly in the first two years of life and is widely regarded as a normal phenomenon. Despite this prevalence, a significant portion of referrals I receive for failure to thrive turn out clearly to be repositioning when I assess the shape of the growth curve. This kind of shift doesn't even take into consideration children such as Michael or my daughter, who may be on their own line far below the curve and are perfectly healthy, nonetheless.

In a similar vein, many children grow in a discontinuous fashion. Rather than a smooth line, if enough measurements are taken, one often sees a stairstep pattern indicating periods of caloric intake without weight gain followed by growth spurts. If these stairsteps are visualized in a small enough window, it can seem as if normal growth suddenly arrests, but if you zoom out, the line smooths. This pattern can be even more pronounced once a child is diagnosed with failure to thrive, because additional monitoring takes more data points closer together in time. These stairsteps can be misinterpreted as the earliest signs of a curve flattening, which could create more concern and more testing, even when unnecessary.

These systematic deficiencies in growth curves are probably more pervasive than is usually taught. My research group at Indiana University School of Medicine collected data from a large cohort of 9,369 children in an Indiana primary care network and tracked the expected weight percentile at 12 months compared with where it was at 1 month. In the paper we published in *JAMA Pediatrics*, we found that nearly half of children fell by one or two lines on the growth curve between the ages of 1 and 12 months. Far from "following the curve" as we might expect, most kids drifted downward. Something significant must be happening between those time points, right? Usually not.



Stephanie Freese

Individualizing growth curves would require a tremendous amount of data as input into a machine learning program. Such models would need to consider many variables, including genetics, medical history, parental sizes and medical histories, the growth rates of similar patients, and previous growth. Forming such models has the potential to help guide overburdened clinicians with flagging patients who really have a problem, but also it would raise serious concerns about data privacy and implicit bias.

How We Make Growth Curves

Growth curves are so limited in their ability to flag potential health problems because of how they are made, a weakness that is difficult to remedy.

Plotting a person's weight and height against a bell curve of measurements from a population has been in use since at least the early 19th century, but not until the 1940s was a healthy population living in ideal standards used to construct curves that look similar to the ones we use today. These growth curves—created in 1946 by public health researchers Harold Stuart of the Harvard School of Public Health and Howard Meredith of the State University of Iowa using a sample of a few hundred healthy, relatively affluent, white children in Iowa—were the dominant reference used by doctors for the next few decades.

The first nationally and internationally recognized growth curves using much larger and more diverse samples were produced in 1977 by the U.S. Centers for Disease Control and Prevention (CDC, then called the Center for Disease Control) and then adopted by the World Health Organization a few years later. These curves were revised in 2000 to include larger and more diverse sample sizes. Regardless of sample size or diversity, all these curves suffer from the same methodological weakness: They are constructed using cross-sectional data. The 1977 CDC curves use roughly 500,000 to 1 million data points in each three-month block, plot the normal distribution of values at each interval, and then use a line-smoothing technique. By plotting the individual measurements for any given point in time, a clinician can see where a patient stands in relation to the population, but that does not clearly demonstrate if the growth rate and trajectory are associated with a medical problem or not. For that, the curves would need to be created using longitudinal data, which is much more challenging to collect.

Of course, the opposite problem to failure to thrive, obesity, has further complicated the definition of "normal" weight, often with significant variation from one location to the next. The latest version of the CDC curves, published in 2020, seeks to address the obesity epidemic more directly by building expanded body mass index (BMI) percentile curves based on updated data through 2016. However, this update continues to further the notion that deviation from normal weight means deviation from health and can cause stigmatization from the health care system. Being healthy is a complex endpoint, and reducing individual well-being down to a single measurement is fraught, as Michael's family will readily tell you. We must vigilantly guard against the tendency to equate "normal" on a curve or a laboratory test with what's "healthy" or "better" for a unique individual.

Improving Growth Curves

At this point, you might think I'm out to get growth curves. But I'm actually a big fan. Growth curves (especially for height) are extremely cost-effective when deployed across large populations. Nutritional deficiencies that can be caught by growth curves have profound impacts on morbidity and mortality decades later, so a tool that can detect malnutrition as a consequence of inadequate caloric intake or chronic disease is valuable. Growth curves can be useful as a component of an overall clinical assessment, because normal growth can rule out many potential problems when families are seeking advice on other symptoms. We absolutely need a tool like growth curves; we just must be conscious of their limitations.

Doctors need access to better models, and training in how to use them. To achieve better models would require longitudinal modeling of very large, diverse populations to detect between-persons differences in within-person change—a broad set of techniques called *growth curve modeling*. There are many statistical approaches, but in essence researchers attempt to fit observed changes over time to a regression model. This kind of method allows comparisons between groups and the creation of curves that more accurately pinpoint where a patient should end up as they grow and gain weight.

An extension of this approach would be to build a predictive model that could be applied to individual patients. Artificial intelligence approaches are likely to provide this solution. Imagine an automated screening tool that tells clinicians, in real time and with some

degree of probabilistic certainty, whether an individual patient has a growth trajectory that indicates the presence of some previously undetected medical problem. This tool could give overloaded clinicians better guidance on when to flag a deviation from the curve as a problem requiring more workup.

If patients meet some threshold in the model, it could trigger the search for a cause while simultaneously limiting unnecessary testing or intervention. AI could synthesize all data in a patient's chart to create an individualized curve, and then provide this guidance in real time to clinicians. For instance, our group has built machine learning models with accuracy as high as 85 percent to predict future risk of obesity using growth chart data and other clinical parameters in the same primary care network we used for our previous study. Although we still need to study this model in the long term, early results are promising that lifestyle interventions can focus on at-risk individuals long before their health is threatened.

These tools are far from being the standard of care, and getting there will require substantial investment of resources and collaboration across health systems. Even when functioning as intended, though, AI models applied to growth curves will not necessarily result in improved health and may generate additional problems. The risk of bias in AI is a systemic problem across every health domain where it has been deployed. Avoiding the stigma associated with, for example, labeling children as obese before they even meet standardized criteria for the diagnosis will require forethought, vigilance, and buy-in from stakeholders across the health system.

Unfortunately, to accomplish more than simple prediction of binary health outcomes with AI is a truly massive informatics task. To build each patient a customized growth curve, models would need to consider many variables, including biological sex, race and ethnicity, geographic location, medical history, and genetics. Furthermore, an interoperable data standard would need to exist nationally and internationally, which opens numerous concerns about data safety and privacy.

All past and present growth curves, regardless of sample size and diversity, suffer from the same methodological weakness: They are constructed using cross-sectional data.

What's more, the introduction of social and demographic variables into such a model raises the specter of implicit bias once again. Even if the model's creators have every intention to generate equitable results and the model is trained on the highest quality data available, its reliance on a system riddled with inequity could easily result in unacceptable bias. In my mind, there is only one solution to this problem: careful, thoughtful, and intentional involvement of the communities that might be affected. The patients who could benefit most and be harmed most by unleashing AI on the health care system need to have a say in how it will be used.

In the end, we will still need to augment the education of clinicians at all levels on the limitations of growth curves. Because we use these tools as screening (or diagnostic) tests, providers must integrate test characteristics such as false positive and false negative rates into their decision-making. When the curves inevitably get better, we will then need to update our clinical education.

I cannot help but suspect the problems we encounter with growth curves stem from other systemic problems in modern clinical care: too much information, too little time, and mounting fragmentation in care. These problems are much harder to fix. Ten years ago, the average patient generated about 80 megabytes of health data per year (that's 25 petabytes per year for just the United States), and the amount of data has only continued to escalate. People in the United States show up to outpatient visits more than a billion times per year, and half of these are to primary care. The true solution is probably to make sweeping changes to the health care system that incentivize more time spent thinking about, examining, and talking with each patient. In the meantime, doctors need to take time to listen to patients; patients should be aware of the limitations of growth curves and ask questions; and researchers and regulatory agencies who study growth should find better ways to present their data.

This individualized care is my greatest hope for children like Michael and my daughter. Vulnerable patients require additional attention to ensure they aren't being over- or underdiagnosed, and to do so we need concerted efforts to improve our tools. For any parent, reassurance that everything is going as it should is of immense value, and we have the data, analytic frameworks, and clinical need to do better.

Bibliography

Bennett, W. E., K. S. Hendrix, R. T. Thompson, A. E. Carroll, and S. M. Downs. 2014. The natural history of weight percentile changes in the first year of life. *JAMA Pediatrics* 168:681–682.

Dugan, T. M., S. Mukhopadhyay, A. Carroll, S. Downs. 2015. Machine learning techniques for prediction of early childhood obesity. *Applied Clinical Informatics* 6:506–520.

NCD Risk Factor Collaboration. 2017. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2,416 population-based measurement studies in 128.9 million children, adolescents, and adults. *Lancet* 390:2627–2642.

Stuart, H. C., and H. V. Meredith. 1946. Use of body measurements in the School Health Program. *American Journal of Public Health and the Nation's Health* 36:1365–1386.

Wood, A. R., et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46:1173–1186.

William E. Bennett, Jr, is an associate professor of pediatrics and a pediatric gastroenterologist, clinical informatician, and clinical researcher at Indiana University School of Medicine. He studies the intersection of clinical informatics and patient engagement research and is especially interested in how the health care system impacts children with neurodevelopmental disabilities. His daughter, Cecilia (age 7) has also provided essential editorial contributions and asks that we include mention of flamingos or the color pink in some way. Email: webjr@iu.edu; Twitter: @WEstusBennett

Cosmos Ex Machina

The universe exists on scales of time and distance that lie entirely outside the range of human experience. It is dominated by two substances—dark matter and dark energy—that we cannot yet create, capture, or even measure in a lab. In the face of our ignorance, relegated as we are to a fleeting moment on one small planet, it may seem an absurd ambition for us to make sense of it all. Yet that is exactly what we cosmologists attempt to do.

We work to combine observations, mathematical models, and computer simulations to retrace the path from the chaos of the Big Bang to the modern universe. Astonishing as it may seem, we are succeeding. Propelled by a succession of ever-more-powerful telescopes, along with modern supercomputers that can perform millions of calculations in a trillionth of a second, we can now provide a detailed account of the growth and development of galaxies over cosmic time.

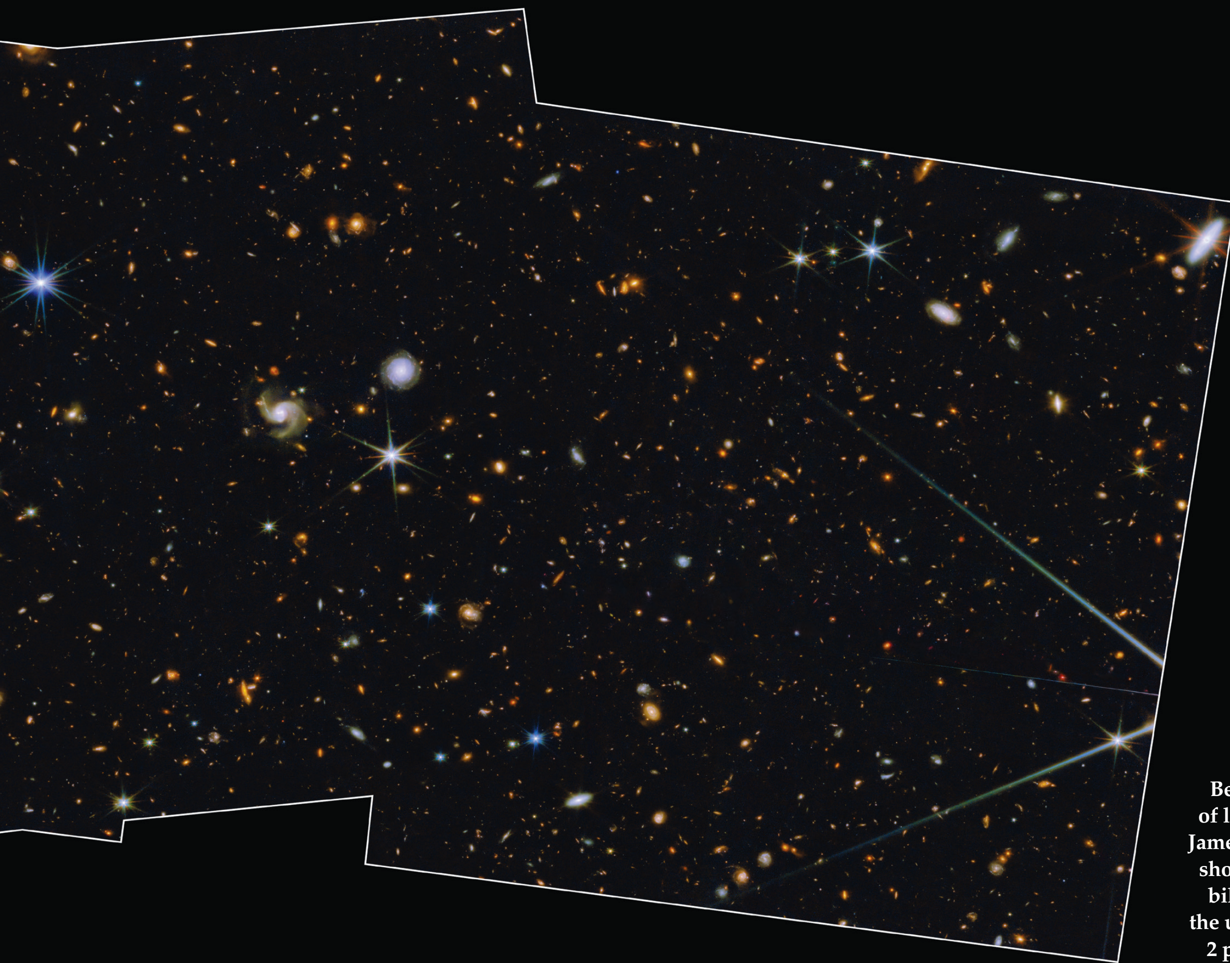
Just a few decades ago, it was not at all clear that we would be able to reach this point. As recently as the 1990s, cosmologists had not yet discovered dark energy, an omnipresent and elusive form of energy that is now recognized as the driver of the accelerating expansion of the universe. Researchers disagreed about the expansion rate of the universe by a factor of two, a discrepancy that provoked bitter arguments at scientific conferences. And astrophysicists could only speculate about when the first galaxies began to form.

Then in the 2010s, the European Space Agency's Planck satellite pinned down the recipe for the modern universe: 68 percent dark energy, 27 percent dark matter, and just 5 percent atomic matter. Planck's high-precision measurements of cosmic radiation also indicated that the universe is 13.8 billion years old, with an error of just 23 million years. Analyses of distant stars and supernovas have determined the expansion rate of the universe to within 10 percent. Most recently, the powerful James Webb Space Telescope (JWST)—which launched in December 2021—has shown us that the era of galaxies stretches back at least to the time when the universe was a mere 320 million years old and was about 1/14th its current size.

JWST is also keeping us humble. Its early observations indicate that galaxies formed earlier and faster than we had expected, in ways that our models did not predict. These findings are forcing us to reexamine our ideas about the earliest generation of stars and galaxies. But out of this current confusion could emerge the next revolution in our understanding of the universe.

Universe in a Box

In some ways, nature has made it remarkably easy to simulate the universe. If we can know the energy content of the universe at any one time (with matter defined in terms of its equivalent energy via $E = mc^2$), we can



Because of the finite speed of light, this image from the James Webb Space Telescope shows galaxies as they were billions of years ago, when the universe was as young as 2 percent of its present age. Many of these early galaxies appear unexpectedly mature.

NASA, ESA, CSA, A. Pagan (STScI), and R. Jansen (ASU)

simply plug those numbers into the equations of general relativity and understand how the energy density of the universe has evolved at all other times. Nature has also provided a blueprint of what the early cosmos looked like. We can observe the *cosmic microwave background*—relic radiation that has traveled to us unimpeded from a time 370,000 years after the Big Bang—to study the highly homogeneous (but not perfectly so) primordial distribution of matter and energy that seeded the galaxies we see today.

With this knowledge, we can predict how dark and atomic matter assembled into collections called *halos*, the formation sites of galaxies. Using the biggest supercomputers and cranking through the equations for gravity, my colleagues and I can evolve simulated mini-universes and see how they change over time. Our models, combined with observations of the real universe, tell us that structure grows hierarchically. Small halos form first, dominated by dark matter; their atomic matter is too sparsely distributed to form stars. The inexorable pull of gravity subsequently brings together many smaller dark halos to form larger ones. It takes time for them to grow massive enough to trigger the formation of galaxies.

Ordinary matter may make up only 5 percent of the universe, but it produces most of its complexity. Protons, neutrons, and electrons interact in varied and complicated ways. The laws that govern atomic matter and its interactions with radiation (and, through gravity, with dark matter) are well known, but the outcomes are hard to predict. Gas cools, condenses, and forms stars; the stars, in turn, inject energy and momentum into ambient gas during their lives and their often-spectacular deaths.

It is essentially impossible to analyze these competing processes by hand. Instead, we feed the equations of gas dynamics into our supercomputer simulations and explore the consequences. We use two complementary classes of simulations to model the universe. The *ensemble technique* attempts to simulate many galaxies from a representative section of the universe, encompassing a volume that is hundreds of millions to billions of light-years across. The *zoom-in technique* places a computational magnifying glass on individual galactic systems and explores these in great detail. Each approach has its benefits and its limitations. The ensemble technique allows us to make predictions about collections of galaxies but can resolve individual systems only fairly coarsely; the zoom-in technique can provide spectacular detail but only on a galaxy-by-galaxy basis.

Although my background lies in conducting big-picture, ensemble studies, I have been captured by the allure of zoom-in simulations, because they allow us to limit our assumptions and pose more specific questions. Can we form individual galaxies that have thin disks and spiral structures like the Milky Way? Can we also form huge,

featureless balls of stars like the biggest known galaxies? Can we track the details of individual star-forming regions and show how thousands of them interact to produce realistic galaxies? It is exhilarating to put the laws of physics into software code, watch the world's biggest supercomputers evaluate these laws trillions of times, and get back the entire history of a galaxy, showing its gas, stars, and dark matter.

Then comes the hard reality check of comparing our computer simulations with astronomical observations. The physical processes that influence galaxy formation are so complex and operate on such a range of scales—from Solar-System–sized disks of hot gas swirling around black holes to galaxy superclusters that are 10 trillion times larger than those black-hole disks—that we cannot fully simulate them all from first principles. We have to make simplifying assumptions about how physics operates, convert those assumptions into rules that we can express in computer code, and then rigorously evaluate how closely our results resemble the actual universe. When our simulations agree with reality, that is good—but surprisingly, it is not good enough to know that the simulation is accurate. Multiple large simulation efforts can match a wide array of observations, even though the simulations often rely on very different physical models. Agreement with observations is merely a necessary condition, not a sufficient one, to claim that a given model is a physically realistic, predictive theory of galaxy formation.

Observations indicate that galaxies formed earlier and faster than we had expected, in ways that our models did not predict.

Often, we learn the most from disagreement. When observations and modeling disagree, we are forced to scrutinize the source of the discrepancy and to consider whether our modeling merely needs minor adjustments (a few more supernovas here, some more cosmic rays there) or if we are missing fundamental aspects of nature.

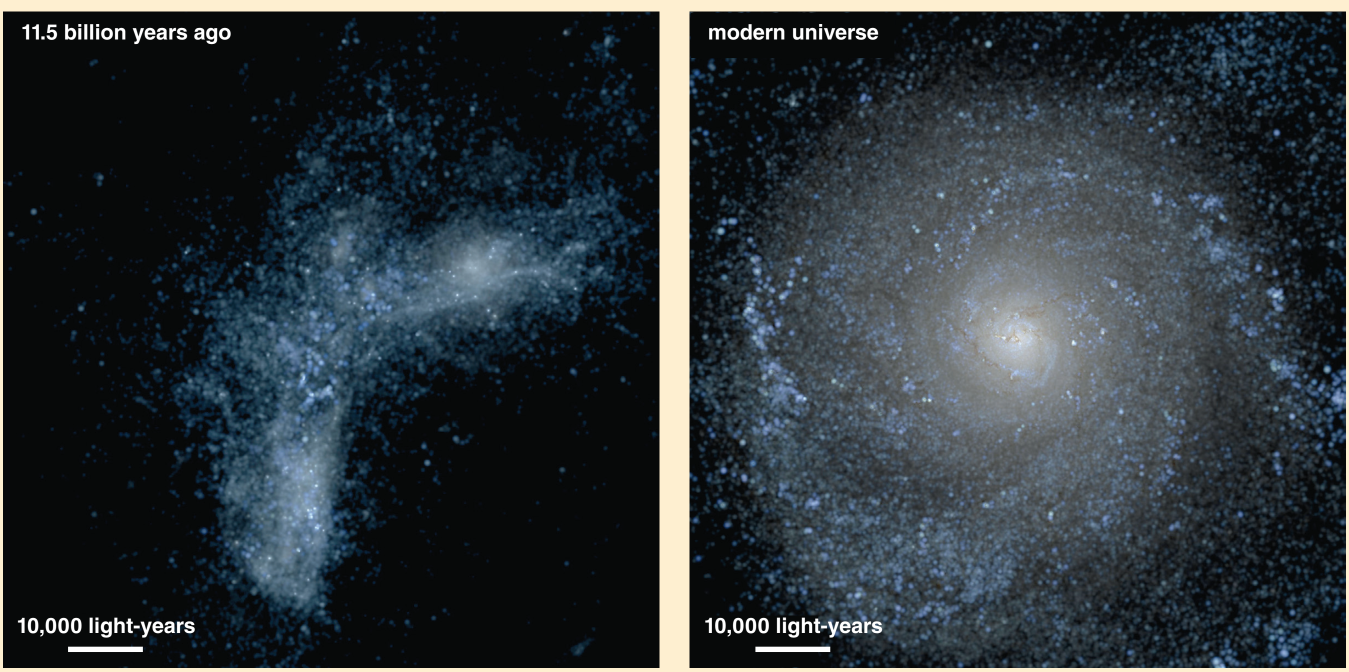
One major disagreement between simulations and observations led to the gradual realization over the past 25 years that black holes have a powerful influence on the properties of massive galaxies. This realization began, as many do, with a seeming contradiction. Massive galaxies at the centers of galaxy clusters are known to contain tremendous reservoirs of gas. All of our knowledge of gas physics said that the gas should be able to cool quickly and collapse, forming a large population of young stars. Observations showed that those galaxies are actually composed predominantly of extremely old stars. Evidently something prevented the gas from turning into stars—but what?

Astronomers knew that the supermassive black holes at the centers of some galaxies can release enormous amounts of energy as they pull in surrounding material. That seemed like a possible mechanism that would heat gas in the galaxies and prevent it from forming stars, but how that process could work was unclear. A breakthrough came when new observations from the Hubble Space Telescope and other facilities demonstrated that supermassive black holes are present in almost all galactic nuclei; meanwhile, theoretical models began to show how energy from those black holes is released and flows through the surrounding galaxy. We now believe that supermassive black holes are indeed crucial to suppressing star formation in massive galaxies, although the details remain a matter of debate. The universe is not so quick to give up its secrets.

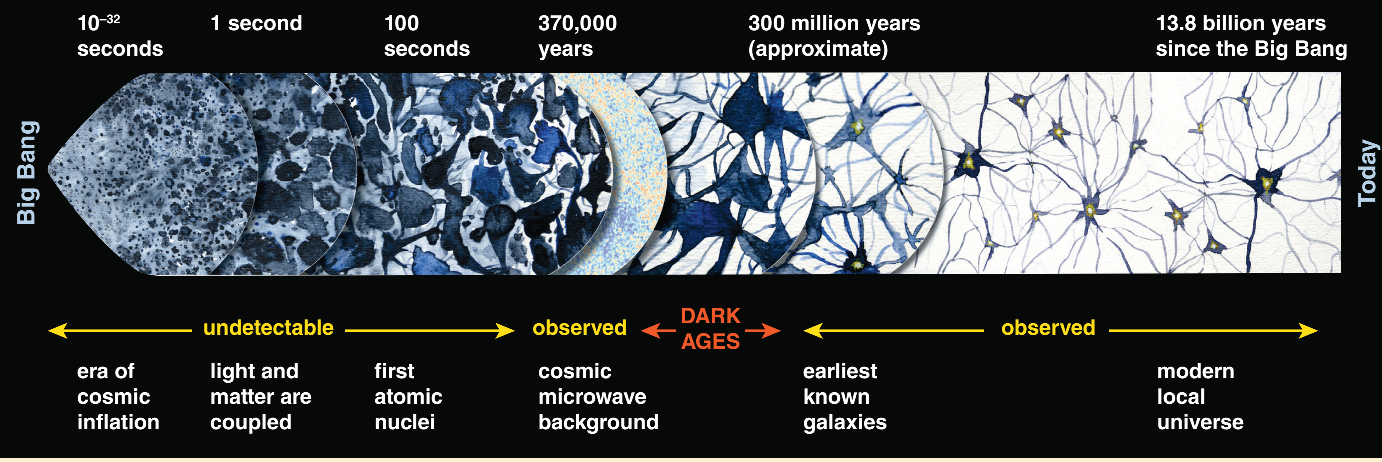
Observing the Past

One of the most exciting capabilities of JWST is its ability to catch light from some of the earliest-forming galaxies.

Computer-generated galaxies from the Feedback in Realistic Environments (FIRE) project allow researchers to rerun the history of the universe. FIRE incorporates detailed information about energy, momentum, mass, and composition to show how formless clouds of gas in the early cosmos (*left*) developed into the well-formed structures seen today (*right*). Such simulations keep getting more accurate, but they still cannot match all the observations.



Jacob Shen/FIRE Collaboration



ESA/Planck, adapted by Barbara Aulicino

This cosmic time line highlights how much of the universe we can observe. It is fundamentally impossible to observe events from before the time of the cosmic microwave background. After that, there was a long gap—the dark ages—before galaxies formed and became visible to us. New telescopes are exploring ever-deeper into the dark ages in search of the earliest stars and galaxies.

Those galaxies are so distant that their radiation has taken billions of years to reach us; over that long journey, the expansion of the universe has stretched energetic light from young stars in those galaxies into longer-wavelength infrared rays that cannot be detected by the Hubble telescope or by ground-based observatories.

JWST was designed to search for the infrared glow of those early stars, and the results are already exceeding expectations. In its first months of operation, the telescope uncovered a startling abundance of well-developed galaxies that appear to have taken shape within the first billion years after the Big Bang. If these observations check out—astronomers are now racing to confirm them—then stars must have formed much faster, and in much greater abundance, in the early universe than previous studies and models indicated.

This news has created an anxious buzz among those of us who study galaxy formation and evolution. Are we misinterpreting the JWST findings? Are we missing something fundamental in our simulations of early galaxy formation? Or do we need to modify the underlying cosmological model on which those simulations are based?

Perhaps JWST has, by chance, observed a highly unusual portion of the sky. Perhaps star formation proceeded differently in the very early universe, when gas clouds did not yet contain the heavy elements that were created by later generations of stars. Perhaps we are actually seeing emissions from black holes and confusing it with starlight. Or, most speculatively, maybe these observations point to a fundamental shortcoming in our cosmological model.

One hypothesis is that there could be a distinct form of dark energy that operated very early in cosmic history (just 50,000 years after the Big Bang), catalyzing the growth of galaxies. As wild as this "early dark energy" scenario may sound, it is not ruled out by observations; in fact, it could help explain a small but notable discrepancy seen between two different ways of measuring the expansion rate of the universe, a problem known as the *Hubble tension.*

Many diverse groups of scientists are investigating these ideas, coming up with models, arguing with one another about both observations and predictions, and disagreeing about the underlying causes. The process may look messy or confusing from the outside, but each of these possibilities is being carefully vetted. Disagreement between our expectations and observations is what drives improved understanding, and sometimes overturns established models. This situation is the cauldron of progress, caught in mid-boil.

We are likely to learn more soon: These surprises come from a relatively small amount of data taken from just the initial JWST observations. Much more data have already been collected and are being analyzed. JWST was so efficient in getting to its observation location (an orbit that keeps it 1 million miles from Earth) that its nominal five-year mission may be extended to 20 years or more. And over the coming decade, many other new observatories—from the European Space Agency's Euclid satellite, to NASA's upcoming Roman Space Telescope, to the Dark Energy Spectroscopic Instrument in Arizona, to the Vera C. Rubin Observatory taking shape in Chile—are set to explore the nature of dark matter, dark energy, and galaxy formation.

There will almost certainly be additional surprises, and we will have to update our models for how galaxies form and evolve, perhaps in major ways. Personally, I am waiting with bated breath for the Rubin Observatory to make sensitive observations of the numerous dwarf satellite galaxies that surround the Milky Way. The number and distribution of those satellites are strongly influenced by the properties of the galactic halos in which they formed, so they should provide insights into the nature of dark matter. The dwarfs are also some of the oldest and least evolved galaxies in the nearby universe; they may be living fossils from the early era of the cosmos that JWST is starting to directly reveal.

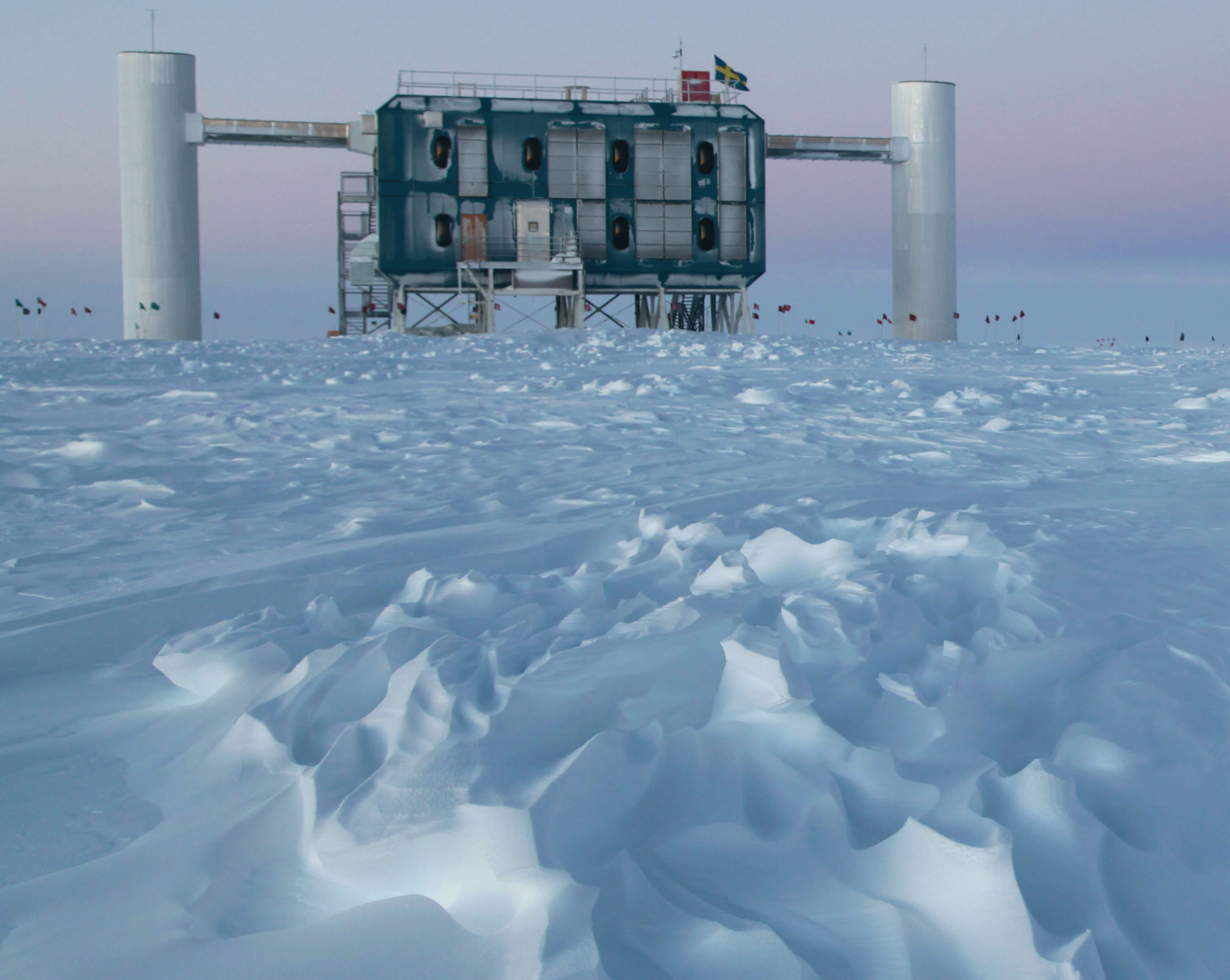
By combining new observations, theoretical models, and computer simulations, we are getting closer than ever to our goal of understanding how the universe works. What will our picture of galaxy formation and cosmology look like in 20 years? I cannot say, and I cannot wait to see.

Mike Boylan-Kolchin is a theoretical astrophysicist at the University of Texas at Austin who works on galaxy formation theory and its interface with cosmology. Recently, his research has focused on near-field cosmology, using detailed studies of nearby galaxies to address a wide variety of questions related to dark matter and galaxy formation physics across cosmic time. His work combines numerical simulations, analytic models, and observations. Email: mbk@astro.as.utexas.edu

A Frozen Window

The IceCube Observatory provides a glimpse of the unseen.

Carlos Argüelles-Delgado



QUICK TAKE

Neutrinos are lightweight particles that rarely interact with other matter, allowing them to travel for millions of light-years across the universe.

IceCube is a telescope at the South Pole that can detect these ghostly neutrinos, enabling scientists to trace their origins in distant galaxies.

Studying how neutrinos change their "flavor" during their journey across the cosmos is helping to test new theories on quantum gravity and the origin of neutrinos' mass.

to the Universe

Every new way of looking at the universe opens previously unknown areas of science. The first optical telescopes allowed astronomers to chart the movement of planets and moons, which helped Isaac Newton formulate his law of universal gravitation. Since then, other kinds of telescopes have enabled us to see invisible forms of light—including infrared, x-rays, and radio waves—that have revealed newborn stars, black holes, and even the faint afterglow of the Big Bang.

Over the past decade, astronomers have begun to see the universe in an entirely new way. This universe is illuminated by tiny subatomic particles called *neutrinos*. Researchers are using these neutrinos to test speculative theories of quantum gravity and to probe the unseen dark matter that makes up around 85 percent of the matter in the universe.

Neutrinos are very lightweight particles, having less than one-millionth the mass of an electron. They are also electrically neutral and nearly inert, interacting with other particles only via the weak nuclear force. On the other hand, neutrinos are extremely abundant. The universe is flooded with neutrinos created in the early universe; many more are constantly being created by nuclear processes or in the collisions of high-energy particles. The closest wellspring is our own Sun, where fusion reactions create a prodigious flood of neutrinos. Looking farther into space, the hot gas swirling around black holes also emits neutrinos, and supernovas create intense bursts of the ghostly particles. Some of these cosmic sources are able to emit supercharged neutrinos that carry more than 100 trillion times as much energy as a photon of visible light.

Because they rarely interact with matter, neutrinos are bold explorers of the universe. Neutrinos produced in a faraway galaxy can traverse millions of light-years unscathed; neutrinos born in the core of a supernova fly right through the remains of the star as if it wasn't even there. These particles fly through you, too. Trillions of them stream through your body every second without having any effect.

This indifference to matter also makes neutrinos exceedingly difficult to study, because most of them leave no footprint as they pass through our instruments. To capture these fleeting particles, we need an absolutely enormous array of detectors, such as those at the IceCube Neutrino Observatory at the South Pole. Then comes a painstaking process of data collection, analysis, and interpretation to discern the traces they leave as they pass. The end result is spectacular, however: My colleagues and I are finally bringing the mysterious neutrino universe into focus.

Across the Universe

IceCube, the world's largest operating neutrino telescope, has been gathering data for more than a decade. Buried about 1.5 kilometers deep in the antarctic ice, it consists of an array of more than 5,000 sensitive light detectors called *digital optical modules* (DOMs) that are strung along cables that have been buried in drill holes. These DOMs are arranged in a roughly hexagonal grid that occupies a volume of approximately 1 cubic kilometer, looking out into approximately 1 billion tons of ultraclear glacial ice.

IceCube's detectors cannot sense neutrinos directly, so they have to look for indirect signs. When a high-energy neutrino occasionally (*very* occasionally) interacts with the ice, it produces charged particles that move faster than the speed of light in that medium. The result is generated light known as *Cherenkov radiation*. Akin to the sonic boom that denotes an airplane traveling faster than the speed of sound, Cherenkov radiation is the visible signature of a passing neutrino. Most of the Cherenkov light is lost as it travels through the antarctic ice, but some of it hits the DOMs, which record precisely when the light arrives. Scientists then use data about the amount of light detected across the array, and the light's exact time of arrival at each DOM, to infer the direction and energy of the initial neutrino (*see figure on page 228*).

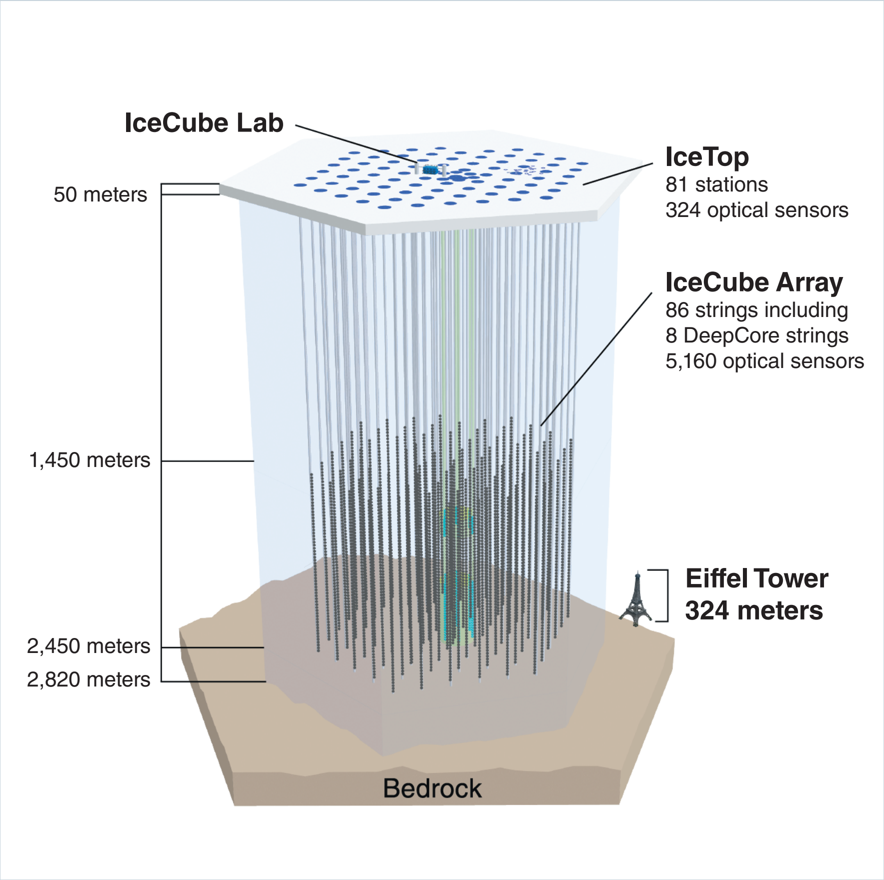
These observations can also reveal what kind of neutrino triggered each flash of light. Neutrinos come in three *flavors*—electron-, muon-, and tau-neutrinos—named after the types of particles they produce when they interact with matter. Thanks to the differences between these three particles, each neutrino flavor leaves a characteristic footprint in our detectors.

For example, energetic electrons can generate a shower of other electrons that each throw off Cherenkov radiation in a revealing cascade. Muons—the heavier cousins of electrons—travel long distances and then disintegrate, leaving extensive tracks in our detectors that are easily spotted. Taus are even heavier but extremely short-lived, producing only small trails between production and disintegration that are very difficult to detect. Only about a dozen individual tau-neutrinos have ever been identified, and it is the least studied of all neutrino flavors.

In 2013, two years after completing the detector array, the IceCube Collaboration announced that it had glimpsed very high-energy neutrinos arriving from distant, unknown sources. It was the first time that such high-energy neutrinos had even been seen coming from outside our Solar System. Unfortunately, we could not track these neutrinos back to a specific point of origin. Although IceCube's detectors can roughly work out which part of the sky a neutrino has come from, the observatory does not have enough resolution to identify an individual source. It's hardly surprising—

The IceCube Neutrino Observatory, based at the Amundsen-Scott South Pole Station, is surrounded by the frozen wastelands of Antarctica. It sits atop a vast array of detectors buried in the ice, which can identify neutrinos from distant cosmic sources. These detectors not only help scientists understand those neutrino sources—such as the roiling clouds of matter around black holes—but also provide a way to test new theories at the cutting edge of fundamental physics.

Sven Lidstrom, IceCube/NSF



IceCube Collaboration

Deep beneath the IceCube Observatory, more than 5,000 digital optical modules (DOMs) are arranged within a cubic kilometer of ultraclear glacial ice. Each DOM watches closely for tiny flashes of light that indicate a passing neutrino.

if you put your thumb up to the sky, some 50 million galaxies may lie behind it. A neutrino arriving on Earth from that general direction could have originated in any one of those galaxies.

But we can draw on other data to help us deduce a neutrino's origins. For example, a neutrino that came from a particular part of the sky could be linked to an event that was observed in some form of light using other telescopes. In 2017, for example, IceCube detected a high-energy neutrino that coincided with a flare of high-energy gamma rays from a source called TXS 0506+056. That object is a type of highly active galaxy known as a *blazar*. It was the first time that a neutrino detection had been pinned to a cosmic source beyond our galaxy.

Recently, researchers have improved the calibration of IceCube's DOM detectors to provide better directional resolution. In 2022, this upgrade enabled the team to trace a high-energy neutrino back to NGC 1068, a galaxy more than 45 million light-years away from us. Like TXS 0506+056, NGC 1068 is an active galaxy.

Although we are still not able to pinpoint the specific objects within those galaxies that generated these neutrinos, other telescopes have shown that both TXS 0506+056 and NGC 1068 contain a supermassive black hole at their center. Those black holes are violently consuming matter, their intense gravitational field drawing a dense and swirling cloud of gas and dust around them. Within these clouds, protons are accelerated to near–light speed. Researchers think that particle collisions in the clouds disgorge the high-energy neutrinos that IceCube has spotted. Thus those neutrinos appear to be messengers arriving right from the precipice of distant black holes.

Over the past decade, astronomers have begun to see the universe in an entirely new way. This universe is illuminated by tiny subatomic particles called *neutrinos*.

Flavor Balance

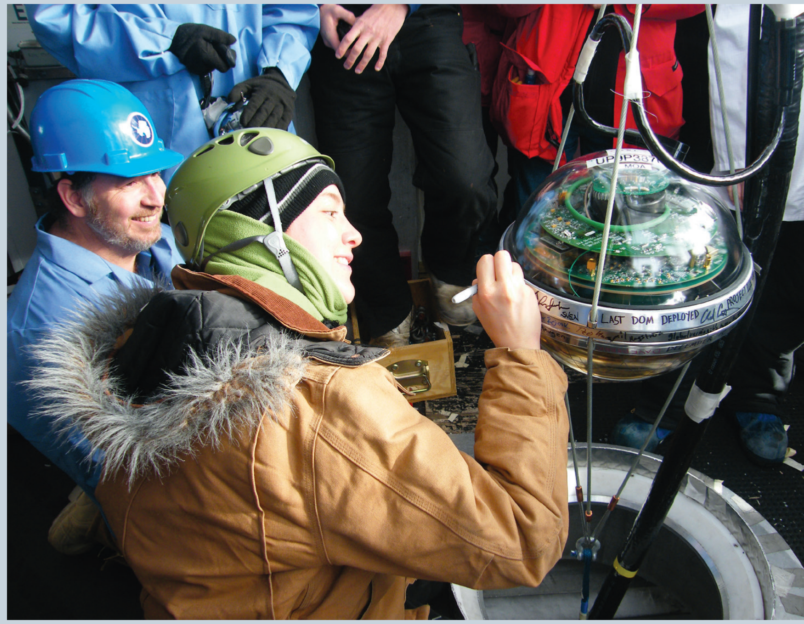
Neutrinos not only carry information about the sources that spawned them; they also bring with them a record of the physical laws that shaped their journey across the cosmos. Researchers working with IceCube are now using the flavor of detected neutrinos to test and improve theories that describe the fundamental principles of the cosmos.

Particle physics experiments on Earth predict that distant cosmic sources should generate about twice as many muon-neutrinos as electron-neutrinos, and very few tau-neutrinos. After the neutrinos are created, another physical process occurs, however. As a neutrino travels through space, it can change from one flavor to another due to a phenomenon known as *neutrino oscillation*. By the time a bunch of neutrinos reach Earth, we expect that oscillations should have produced an equal mix of flavors.

Any small deviations from an equal ratio of neutrino flavors would suggest that the particles were affected by unknown processes during their journey across the cosmos. By studying the flavor ratios of neutrinos that arrive at IceCube, we can potentially see the effects of these processes and test theories that go beyond our current understanding of physics. In particular, I have collaborated with Teppei Katori of King's College London and Jordi Salvadó of the University of Barcelona to show that the ratio of neutrino flavors offers one of the most sensitive tests of quantum gravity.

Quantum gravity seeks to merge two of the most successful theories in physics: quantum mechanics and Albert Einstein's general relativity. Quantum mechanics describes the dynamics of very small objects, such as a photon or an electron orbiting the nucleus of an atom, whereas general relativity is a theory of gravity that explains the

DESY



Gary Hill, IceCube/NSF

These DOMs (*left*) contain light detectors sealed inside transparent spheres. The detectors are sensitive enough to register evidence of passing neutrinos. To build the detector array, the IceCube team used jets of hot water to drill 86 boreholes in the antarctic ice, each 60 centimeters wide and 2.5 kilometers deep. A researcher (*above*) signs a DOM just before it is carefully lowered into the depths.

large-scale evolution of the universe. Unfortunately, these theories cannot both be completely true. In some circumstances, they give wildly inconsistent or nonsensical answers.

Quantum gravity aims to overcome these problems by forming a bridge between the two theories. The effects of quantum gravity will be incredibly subtle in most situations, and only become significant at extremely small distances or at extremely high energies. One potential signature of quantum gravity involves the breakdown of a fundamental symmetry in the universe. This symmetry rests on the idea that there is no preferred direction in the universe—in other words, the cosmos has no label showing an arrow and the word "up."

If this symmetry is broken, it means that there are discernible elements in space-time—glitches in the very fabric of the universe, you might say—that do indeed point in a particular direction. And that's where neutrinos come in. As they travel from their faraway sources to Earth, neutrinos can interact with these space-time elements in ways that imprint distinct signatures on the neutrino flavor states. For example, some quantum gravity theories suggest that the interactions can switch off neutrino oscillations, forcing the neutrino to maintain the same flavor from source to detection. In some scenarios, this phenomenon could make all of the neutrinos from a source arrive on Earth as a single flavor, rather than the three-way mix we would otherwise expect.

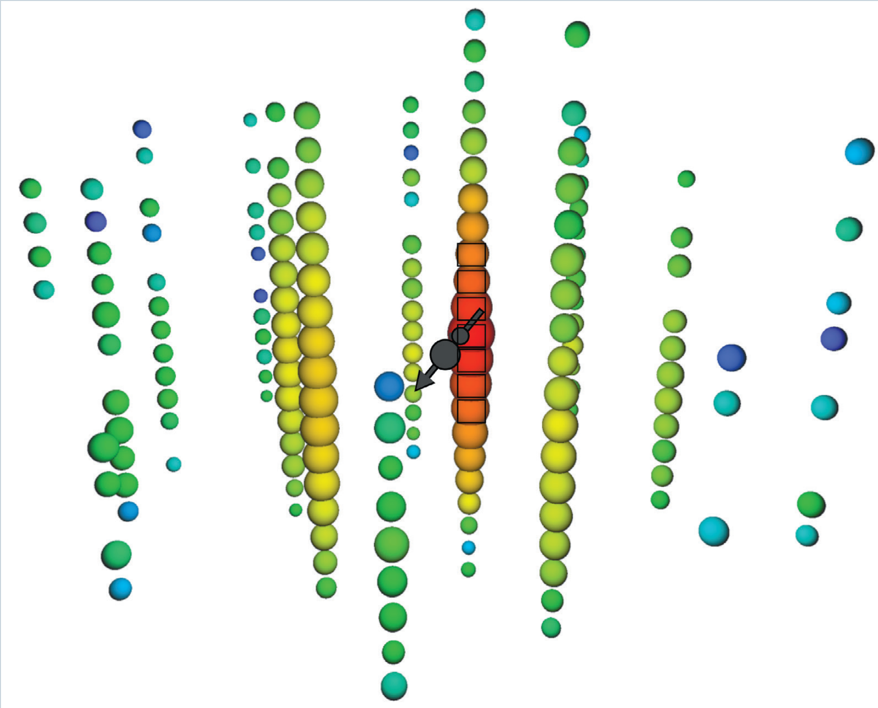
To test these far-reaching ideas, we first needed to prove that IceCube can detect all flavors of neutrinos—even the super-obscure tau-neutrinos. That effort pushed our observatory to its limits. Last year, I was part of the team that announced the first detection of a tau-neutrino from a cosmic source. Because the neutrino almost certainly did not start out with its tau flavor, our detection confirmed that it must have undergone neutrino oscillation during its vast journey.

Researchers working with IceCube are now using the flavor of detected neutrinos to test and improve theories about the fundamental principles of the cosmos.

We were able to identify this tau-neutrino by its unique light signal picked up by our IceCube detectors. When a tau-neutrino hits the ice, it produces a very short-lived tau particle and an accompanying flash of light. In about 100 billionths of a second, the tau disintegrates into other particles, releasing more light. So the signature of a tau-neutrino is a distinctive double flash: one signaling the production of the tau particle, followed by a second one that shows its disintegration. For very high-energy neutrinos, these two emissions might be only about 50 meters apart in the ice surrounding IceCube, the average distance a tau with a million times the energy of a proton can travel during its fleeting existence.

To look for these signatures, I worked with an international team to meticulously reanalyze IceCube data collected between 2010 and 2017. This tau-neutrino hunt was led by Juliana Stachurska, currently a postdoctoral researcher at the Massachusetts Institute of Technology.

While our team was at an IceCube Collaboration meeting in Atlanta, we



IceCube Collaboration, 2022, *European Physical Journal C*

IceCube researchers recently identified an elusive tau-neutrino by its characteristic signature. Each colored circle represents a burst of light detected by the IceCube sensors. The color shows when the light was detected (red is earlier, blue later**), and the circle's size shows how much light was detected. The two gray circles and arrow show the beginning and end of the tau's brief life, and its trajectory.**

went over our initial checks and found some confusing results. We all decamped to a coffee shop to figure out what was going on. Picture the scene: A group of young scientists huddling in the coffee shop, checking many lines of their computer code and poring over figures, desperately searching for a rare subatomic particle from deep space. Eventually, we looked at the raw data itself. We immediately saw that a pair of light sensors had captured two consecutive light emissions just 17 meters apart, a clear sign of a tau-neutrino.

When people discover a new neutrino, they usually get to name it—that is how rare these detections are. We called our precious tau-neutrino *Double Double*, partly because of its characteristic light signature, but also in honor of the classic double-cream-and-sugar coffee sold by the Canadian restaurant chain Tim Hortons. We were drinking a lot of coffee back then.

Our observation of the Double Double tau-neutrino is already helping to test quantum gravity theories. For now, the data show no sign of the space-time defects predicted by quantum gravity theories. That's disappointing, but instructive. Every result will help physicists to refute and refine their ideas about how to unite quantum mechanics and general relativity.

A Mass of Data

Although IceCube has confirmed only two specific neutrino sources—the active galaxies TXS 0506+056 and NGC 1068—it has spotted hundreds of other high-energy neutrinos. By combing through these data, we should be able to confirm other sources that will give researchers more ways to study neutrino physics. An identified source provides a precise distance between where the neutrino was produced and where it was detected. Because neutrino oscillations depend on this distance, combining a known source location with information about the flavors and energy distribution of the neutrinos will fill in crucial pieces of the neutrino puzzle.

For example, researchers led by Kiara Carloni, a graduate student in my research group at Harvard University, recently proposed a way to use distant neutrino sources to uncover the nature of a neutrino's mass, which is one of the biggest mysteries of contemporary physics. For a long time, physicists wondered if neutrinos had any mass at all. Previous experiments found that neutrinos have a mass that is nonzero, but one that is very, very small—so small that it cannot be measured in laboratories. The origin of this mass is unclear. Other particles, such as electrons or protons, gain their mass through interactions with the so-called Higgs field, embodied in the Higgs boson that was discovered in 2012 by the Large Hadron Collider at European Council for Nuclear Research (CERN), near Geneva. But the extreme smallness of a neutrino's mass suggests that some other mechanism is at play.

José Valle at the University of Valencia in Spain posits that neutrinos gain their mass by partnering with so-called "sterile" neutrinos, which never directly interact with any other particles. These sterile neutrinos could be produced by the same kind of neutrino oscillations that enable the particles to change flavor. But even by the standards of our work at IceCube, finding a sterile neutrino would be quite a feat. In theory, an average neutrino would have to travel truly enormous distances to stand a chance of becoming sterile. Although we have seen hints of similar particles in our particle accelerators, Earth-bound experiments are simply not big enough to capture sterile neutrinos.

The good news is that neutrinos from distant galaxies and other cosmic sources have the right energy, and travel for sufficiently long distances, to trigger their conversion to a sterile form. Carloni's work suggests that when this conversion happens, we should see an absence of neutrinos at a particular energy, because they no longer interact with our detectors once they have become sterile. In other words, we can't detect the presence of sterile neutrinos, but we might be able to detect the absence of the neutrinos that sired them.

Before we can perform these tests, we will need to identify many more neutrino sources. Fortunately, other observatories are poised to join IceCube in exploring the neutrino universe. The Cubic Kilometre Neutrino Telescope (KM3NeT) is being built at the bottom of the Mediterranean Sea, while the Baikal Deep Underwater Neutrino Telescope (Baikal-GVD) is taking shape in Lake Baikal in Russia. The first phase of Baikal-GVD was completed in 2021; KM3NeT is scheduled to become operational in 2025. Another venture called the Pacific Ocean Neutrino Experiment

(P-ONE) could be constructed toward the end of the decade off the coast of Vancouver, Canada.

All three of these new neutrino observatories operate on similar principles as IceCube, but they use water instead of ice as the detector medium.

It is astonishing that these kilometer-scale telescopes can turn rock, water, and ice into sophisticated eyes that look deep into the invisible neutrino universe.

Physicists expect that these water-based neutrino telescopes will have a better directional resolution than IceCube, boosting our chances of pinpointing specific objects as the sources of cosmic neutrinos.

We may also gain more insights from proposed neutrino observatories that would use large mountains as their detector medium, relying on solid rock to trigger the neutrino interaction. For example, there are plans afoot to build the Tau Air Shower Mountain-Based Observatory (TAMBO) in one of the deepest high-altitude canyons in the world, up in the Peruvian Andes. TAMBO would nicely complement IceCube and the other neutrino telescopes, because it would be a specialized tau-neutrino detector.

It is astonishing that these kilometer-scale telescopes can turn rock, water, and ice into sophisticated eyes that look deep into the invisible neutrino universe. After a decade of remarkable progress at IceCube, it's clear that the era of neutrino astronomy is only just beginning.

Bibliography

Argüelles, C. A., T. Katori, and J. Salvado. 2015. Effect of new physics in astrophysical neutrino flavor. *Physical Review Letters* 115:161303.

Carloni, K., et al. 2022. Probing pseudo-Dirac neutrinos with astrophysical sources at IceCube. Published online. *arXiv*:2212.00737.

IceCube Collaboration. 2013. Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science* 342:1242856.

IceCube Collaboration. 2018. Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science* 361:147–151.

IceCube Collaboration. 2018. Neutrino interferometry for high-precision tests of Lorentz symmetry with IceCube. *Nature Physics* 14:961–966.

IceCube Collaboration. 2021. IceCube data for neutrino point-source searches years 2008–2018. Published online. arXiv:2101.09836.

IceCube Collaboration. 2021. IceCube high-energy starting event sample: description and flux characterization with 7.5 years of data. *Physical Review D* 104:022002.

IceCube Collaboration. 2022. Evidence for neutrino emission from the nearby active galaxy NGC 1068. *Science* 378:538–554.

IceCube Collaboration. 2022. Detection of astrophysical tau neutrino candidates in IceCube. *European Physical Journal C* 82:1031.

Carlos Argüelles-Delgado is an assistant professor at Harvard University, where his research focuses on searching for new physics with high-energy neutrino data from the IceCube Neutrino Observatory. Email: carguelles@fas.harvard.edu

Building Knowledge

Henry Petroski

Simulation and modeling are virtually synonymous with engineering, for an engineer can hardly proceed with the analytical design of a machine, system, or structure without having at least an image of it in mind. That idea may conceivably promise to be the greatest thing since sliced bread, but it may have to forever remain at best a half-baked, private thought if it cannot be communicated to someone who can engineer it into something concrete. The typical first step in doing so is to simulate or model the idea through images that give it boundaries at which it will interact with the system into which it will be built. In the case of a structure such as a building or bridge, engineers must translate the originator's idea into a form that others can grasp unambiguously and thus test analytically the structure's strength, stability, and sufficiency for fixing the problem that the idea aims to solve.

Courtesy of Linda Hall Library of Science, Engineering, and Technology

Before Galileo Galilei analyzed the cantilever beam with the same scientific rigor he had the heavens, designing a physical structure was largely a matter of craft tradition rooted in trial and error. As late as the 17th century, what worked was imitated; what did not was abandoned. Galileo distinguished himself by asking and pursuing in a rational manner the question of why some structures failed. At the time, it was well known that ships built in a geometrically similar way to ones that were successfully sailing the seas would themselves be successful—up to a point. It was natural to build larger and larger ships based on previously successful designs, but when the largest ones failed upon being launched, there appeared to be a limit to size. But why?

Galileo was able to explain why by changing the common conception of a ship. Rather than considering the entire structure, Galileo modeled the vessel at port as a simple beam supported only at its ends. When the ship leaving

Galileo Galilei's mental modeling transformed a cantilever beam into a lever, as illustrated in his 1638 treatise *Dialogues Concerning Two New Sciences*. Recontextualizing established ideas can lead to the development of new solutions to old problems.

Henry Petroski is the A. S. Vesic Distinguished Professor Emeritus of Civil Engineering at Duke University. His latest book is Force: What It Means to Push and Pull, Slip and Grip, Start and Stop *(Yale University Press, 2022). Email: petroski@duke.edu*

the port is pushed into the water bow-first, there is an instant when one end is still on land and the other is floating in the water. This observation provided Galileo with the structural model of a beam that he could analyze. Galileo extended the problem further by mentally breaking the beam in two: Rather than one simply supported beam extending from the bow, he thought of them as a pair of cantilevered beams attached back-to-back as mirror images of each other, one from the bow and one from the stern. This mental modeling led to what has come to be known as *Galileo's problem*.

Galileo further abstracted the problem by modeling the interaction between the end of the beam and the wall in which it was embedded with a distribution of forces representing the effect of the wall on the beam. How he envisioned this problem was the crucial step in his analysis. By analogy—a common tool in simulation and modeling—a strand of dry spaghetti clutched at one end in the left hand and pulled down at the other end by the right hand bends and then breaks where the rod of pasta meets the clutching hand. The same concept applies to cantilevered beams, as illustrated by a famous engraving in Galileo's 1638 treatise *Dialogues Concerning Two New Sciences* (*see illustration on page 232*). In Galileo's model, as the weight forces the exposed part of the beam to rotate about its base (about an axis through point B and parallel to the bottom of the square-hewn beam), the tenacity of the wood resists that rotation. In other words, he saw the beam as a canted (bent) lever, with the effort coming from the rock hanging from the free end and the resistance coming from the embedded end. In a ship, where the beam is supported at both ends, the full beam would break at its midpoint.

Although Galileo did not use equations in the *Dialogues*, his narrative calculations of balancing the turning effect of the hanging rock with the resisting effect of the embedded end were straightforward and resulted in the verbal equivalent of the mathematical formula $W = Sbh^2/2L$. This equation, which is among the most basic calculations that engineering students tend to learn by heart, calculates the weight (W) that would cause a cantilever beam of a specific length (L), base (b), height (h), and strength (S) to break.

A number of practical engineering conclusions can be drawn from the mathematical model of the physical beam. Its breaking point is proportional to the strength of the material, the width of the beam, and the *square* of the height of the beam. Furthermore, the breaking force for a cantilever is inversely proportional to the length of the beam, which is consistent with our experience—and that of shipbuilders of Galileo's time—that shorter is stronger than longer for beams of the same cross-sectional dimensions. The factor 2 in what may be called Galileo's formula comes from the mathematical analysis, and it held the key to discovering that the formula was wrong!

Even a genius like Galileo could get caught up in looking at his model so closely that he seemed to have forgotten that the final analysis had to be consistent with reality. When the formula began to be employed by engineers to design such things as a piping system to supply the fountains of Versailles, the formula sometimes led to failure. It took almost a century to understand why Galileo's formula was not proving reliable, to correct it, and to disseminate a new one. Galileo's mistake was in assuming the forces of adhesion were uniformly tensile, when in fact they follow a triangular profile from compression at the bottom of the beam to tension at the top. Balancing these effects changes the two to a six in the denominator of the equation. In other words, Galileo's formula was predicting beams to be three times as strong as they really were.

Using computer models can be fraught with danger, especially if the user has not been exposed to the theoretical foundations on which the model is based.

Foundational Models

The use of Galileo's formula in design and analysis led to increasingly more sophisticated models for the elastic beam, including those for two- and three-dimensional models of elastic plates and shells. As an undergraduate student in mechanical engineering at Manhattan College, I and all engineering majors at the time were instructed in models and methods that had evolved from Galileo's work; as a graduate teaching assistant, fellow, and instructor in theoretical and applied mechanics at the University of Illinois in the 1960s, I taught in the spirit of Galileo's analytical technique, but I employed long-corrected models.

My formal education took place during a time when analytical models were increasingly complemented—and in some cases replaced—by computational ones. At Manhattan, I learned about analog computing, in which electrical circuits were used to model mechanical systems by exploiting the correspondences between voltage and applied force, resistance and a spring, capacitance and mass. The simplest of such analog models can represent such mechanical phenomena as a vibrating body or a shock absorber.

The postwar years saw the rapid development of digital computers, in which structural mechanics problems began to be solved by techniques based on what became known as *finite-element models*. In these models, a structure such as a building or bridge is discretized into interconnected simple beams whose combined behavior simulate that of the extended structure. In graduate school, I had my introduction to digital computers, and later at Argonne National Laboratory I collaborated with structural analysts using advanced *finite-element codes* (as generic digital models were often called) to solve problems in nuclear power reactor fuel subassemblies, containment vessels, and piping systems. In particular, I applied principles of mechanics to problems involving cracks and their propagation in such components.

Using computer models can be fraught with danger, especially if the user has not been exposed to the theoretical foundations on which the model is based and does not have a feel for the actual problem involved. In 1980, based largely on my work in various areas of mechanics, I was offered a position in the civil engineering department at Duke University. One semester, I taught a course in classical elasticity, using the third edition of Stephen Timoshenko's classic *Theory of Elasticity*, which was first published in 1934. One of the homework problems I assigned to the students asked them to determine the effect of a small, circular hole in an elas-

University of Washington Libraries. Special Collections, UW 41888

Following the collapse of the Tacoma Narrows Bridge in 1940, engineers at the University of Washington Structural Research Laboratory built a model to study why the structure failed. The 1:150 scale model was inside a wind tunnel to simulate the forces that pushed on the bridge.

tic strip under uniform tension. The mathematical model developed in the textbook boiled down to a differential equation whose solution could be expressed in a polynomial, which is how most students in the class solved the problem, getting the correct answer that the hole magnified the stress by a factor of exactly 3. One student, who had taken leave from his job in the aerospace industry to earn a master's degree, insisted that the answer was something like 2.3456 . . . When asked why, he described how he had used a versatile computer code that was heavily relied upon within his company. After reluctantly agreeing to refine the mesh of his digital model and rerun it a few times, he reported that the output was converging to 3.

Other ways in which models can be misleading at best, and downright wrong at worst, is by using them beyond their intended application. For example, all mathematical and computational models are based on assumptions regarding the material of which a structure is made. However, the nature of much design and analysis is to push the envelope in employing new materials in new ways, which may mean under conditions that require knowledge of material properties at temperatures beyond which they are known and what is embedded within the digital model. Relying on results delivered by a model used beyond its implicit limitations can lead to disaster.

Before the mid-20th century, structures such as bridges were modeled as interconnected beams, plates, and columns and were designed using hand calculations assisted by the (analog) slide rule and what was essentially an electromechanical adding machine. This method resulted in many successful bridges, which in turn served as models for even more ambitious structures. The Tacoma Narrows Bridge was designed in the 1930s as a natural extension of the prevailing model of a suspension bridge: one with a long, slender, unstiffened span. It proved to be modeled beyond experience with the extremes of those parameters and collapsed in the wind. Subsequent analysis of the failure involved physical scale models of the bridge, but those models could capture only gross phenomena and were subject to the criticism that they could not possibly represent the actual bridge, which may have had imperfections of materials and construction that were not incorporated in any physical scale model.

By the 1970s, engineering design offices housed two cultures: younger engineers whose instinct was to use digital models, and the older engineers who grew up thinking analogically.

The 1960s were a transition decade between a generation of engineers who thought and designed analogically and those who did so digitally, and fluency in the computer language FORTRAN became a shibboleth. By the 1970s, engineering design offices housed two cultures: a growing cadre of younger engineers whose instinct was to use prepackaged digital models, and the older, supervising engineers who grew up thinking analogically. Both sides were fully aware of the diverging cultures and of the limitations of models both analog and digital. On several occasions over lunch, engineers of such record-setting structures as the Twin Towers of the World Trade Center and the Burj Khalifa related to me their approach to supervising young engineers wedded to computational models. They asked young engineers to compute the maximum horizontal movement of a super-tall building in the wind. The older engineer then took out his pencil, paper, and slide rule and modeled the building as a vertical cantilever loaded by a horizontal force at its end. If the two calculations were within an order of magnitude of each other, the experienced engineer gained some confidence in the digital model. If not, the inexperienced engineer was sent back to the workstation.

One of the greatest pitfalls in using models in engineering or science is the tendency to confuse the model with what it is supposed to represent. Richard Feynman wrote that the first principle of "scientific integrity" is "you must not fool yourself—and you are the easiest person to fool." Nothing more relevant might be said of the use of models and simulation.

Selected Bibliography

Bush, V. 1945. *Science—the Endless Frontier*. Washington, D.C.: Government Printing Office.

Feynman, R. P. 1985. *Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character*. New York: W. W. Norton.

Petroski, H. 1994. *Design Paradigms: Case Histories of Error and Judgment in Engineering*. New York: Cambridge University Press.

Timoshenko, S. P., and J. N. Goodier. 1951. *Theory of Elasticity*, 3rd ed. New York: McGraw Hill.

The Early Human Social Network

Courtesy of Cecilia Padilla-Iglesias

Paleoanthropologists have long established that the earliest humans came from Africa, but new discoveries are complicating the narrative of exactly where in Africa they originated. Cecilia Padilla-Iglesias is melding environmental data and studies of modern Central African hunter-gatherer populations to learn about ancient humans and how these early populations interacted. Her models have indicated that the first people lived in communities that were far more connected than previously thought, sharing both genetic material and cultural traditions. Padilla-Iglesias, who is a doctoral candidate in evolutionary anthropology at the University of Zürich, combines genetic, archaeological, ecological, and ethnographic techniques to better understand the processes that shape human diversity. She spoke with special issue editor Corey S. Powell about her research. This interview has been edited for length and clarity.

Your work diverges from the idea that modern humans emerged from a single population in East Africa. At what point did you start to think that there were details that didn't match the conventional models?

There is a long tradition of paleoanthropology in East Africa. We know that some of the oldest human fossils were found there, so we must have come from there and left. But in the past 10 to 15 years, very old, human-like things were found in Southern Africa and in Morocco. We saw that some of the oldest genetic lineages were from populations that now live in other parts of Africa. Some researchers started developing a theoretical model in which humans came from more than one place in Africa.

As I was reading this literature, I wondered whether the Central African hunter-gatherers I worked with fit this model. If there was one place that could connect Morocco with Southern Africa and East Africa, it was Central Africa, and I also knew that hunter-gatherers had networks created for mobility.

I started looking at the genetic evidence and environmental models and asking, Where do contemporary Central African hunter-gatherers live? Where are the fossils, the few that we've found in Central Africa? Under which environmental conditions? Then I created a model using data on contemporary hunter-gatherers from Central Africa to try to predict where we've found fossils and genetic divergences in the past.

I found that the moments in which my environmental model predicted fragmentation coincided with points where I could see genetic splits between populations. I'm also doing analysis that can trace the evolution of certain words and objects—almost doing a gene tree, but with objects.

Even if these populations now speak languages that are unrelated, it does seem that these people had a very deep past in which they were connected to one another. They were exchanging genetic material, and I'm finding that they exchanged culturally with one another as well. And if the whole stretch of Central Africa was exchanging genes and culture, they must have been also exchanging genes and culture with people a bit further north, a bit further south, and to either side.

The idea of *connectivity networks* is very important in your research. Do genetic connectivity and cultural connectivity go hand in hand?

Not necessarily. In order to have a diverse set of solutions to complex problems, two things are important: innovation and collaboration. When I talk about connectivity being the key to genetic and cultural diversity and an important element of human evolution, it's the idea that different populations living in different environments that come into contact every once in a while are able to optimize these two processes. The separate populations are each innovating, and when they come together and exchange those ideas, they're able to get the best of what the other one has.

What is the role of *partial connectivity* in this process?

If you have a population that's fully connected from the beginning, everybody would get stuck on the idea that came first or that was from the most prestigious person. Partial connectivity allows populations to explore and build on a set of solutions to build a greater repertoire.

For example, you might have a population that has optimized fishing methods and another one that has great antelope hunting techniques. You can recombine and even trade these ideas. And with genes it happens in a similar way. You're able to access a greater gene pool, even if you're adapted to a local environment.

We conducted a study in which I compared the bits of DNA that I knew had been exchanged between Central African hunter-gatherers with the structural diversity of their musical instruments and of their foraging tools. These two sets of objects have completely different purposes; the musical instruments might serve as identity markers, whereas the tools might be specific to a foraging niche. When I did a relation test between the structure of genetic diversity and the structure of musical instrument diversity, the two matched almost perfectly—populations that were exchanging genetically were also exchanging musically. But for the foraging tools, there was no relationship at all. It just may not make sense to use the tools that someone in a different environment is using.

You are describing deeper and wider levels of connectivity than what's commonly represented in the scientific literature. Are there barriers that have kept people from taking this broader view?

Traditionally ethnographers who worked with hunter-gatherers would

go to one place and spend a lot of time there. They would live with a group for a couple of months and note everything about rituals and about the physical appearance of a population. But nobody would spend long enough to understand the big aspects of mobility, of how far people are moving.

On the other hand, archaeologists looking at comparative pottery analysis and ochre transport have shown that people were using nonlocal materials for a very long time. We've found shell beads in the middle of the continent where there are no shells. It must be that people were trading. Perhaps we've overlooked the scale.

There is nonanecdotal evidence that big networks were in place. Maybe it's not just a consequence of people being friendly or moving around. Maybe those networks were the reason people moved around, to maintain them. Maybe they were way bigger than we've ever thought.

Your studies indicate that foraging is central to hunter-gatherer social networks. Why is foraging so important?
Foraging is an aspect of hunter-gatherer social life that's been studied a lot, traditionally, from an anthropological perspective, because it's a behavior that separates humans from, say, great apes. If chimpanzees find food in the forest, they'll just grab and consume it where they find it. When night comes they'll build a sleeping nest where they are and then abandon it.

Researchers working with hunter-gatherers from an archaeological perspective found that humans didn't just consume food when and where they found it. They would take it back to a central place or a camp where they would share it with others and consume it together. This practice is a huge aspect of hunter-gatherer social life because it allows for the division of labor. You have people who remain at the camp and others who forage, and then the food brought back is for everyone. It's this element of cooperation that, as we often say in anthropology, separates humans from other members of our family tree. It allows people to survive even when they don't forage. One day I get yours and one day you get mine. And it allows people to tell stories and talk about one another.

An interesting aspect my team is exploring is not just the implications of going back to a single place or not, but variations within that spectrum. We know that hunter-gatherer societies around the world vary in how much time they spend in these central places, and how far away from them they're willing to travel to forage. It depends a lot on ecological parameters. If there's plenty of forage close to you, you won't go so far away. If there's less, you will deplete the things that are close by faster, and therefore you potentially need to move your home base earlier in order to not starve.

How do you put your ideas about social networks to the test?
Our team built a model to ask, What are the implications of this changing behavior for early human social networks compared with those of other animals, such as great apes? And, if we find differences in the network properties, what would be the consequences for cultural transmission, which is one thing we know that human societies are exceptionally good at doing?

"If the whole stretch of Central Africa was exchanging genes and culture, they must have been also exchanging genes and culture with people a bit further north, a bit further south, and to either side."

In our model, people move based on resources. Then we calculated the efficiency of the resulting network. We found that the hunter-gatherer way of moving, compared with the great ape way of moving, led to networks that revealed signs of being partially connected. They had very densely connected nuclei according to what we would observe of camps or groups of camps, and they would be embedded in large regional networks.

We also ran a *contagion simulation*, which is a method that comes from epidemiology that assesses how fast and how far a new innovation or idea travels in a network. We found, again, that the hunter-gatherer networks were particularly efficient at transmitting these new innovations quickly in a way that could reach the network nodes. It's almost as if they were adapted to be efficient at diffusing advantageous information across a network. Then we played around with what happens when the environment is more heterogeneous or more homogeneous. We found that—as we would expect—a more heterogeneous, rich environment would facilitate this information transmission and these network efficiencies.

How have computer models influenced your ideas about human development?
The great thing about computer models is that you can explicitly test mechanisms. If I just rely on a statistical model, I could say: "I think that because this river was dry 7,000 years ago, these people stopped exchanging genes," and then compare gene exchange before and after the river dried. But it could have been the river, or it could have been seven other things. With computational modeling, you can create a world in which the only thing that changes is that river. If you have sufficient processes in place that mimic the real world, you can see how big an impact the river drying would have had. Then you can test your findings against your data.

For example, our lab simulated something simple, such as the development of a plant medicine that involved innovating with different plants and recombining them. We simulated that process in different types of networks, each with the same number of nodes and edges but different structures. One network was fully connected, and in another every edge and every node could be connected or not in a random fashion.

We also ran the simulation on real networks collected from hunter-gatherers. We could compare directly—not varying anything else—how the network structure affected the speed at which a group would reach the right medicine. The hunter-gatherer version was much more efficient than a random configuration or a configuration that was fully connected. There's something mechanistic in the structure of this network that makes it efficient at discovering a complex solution through recombination. For my work, especially when you add the spatial component, it opens an entire new realm of possibilities.

A companion podcast is available online at americanscientist.org.

Farming and the Risk of Flash Droughts

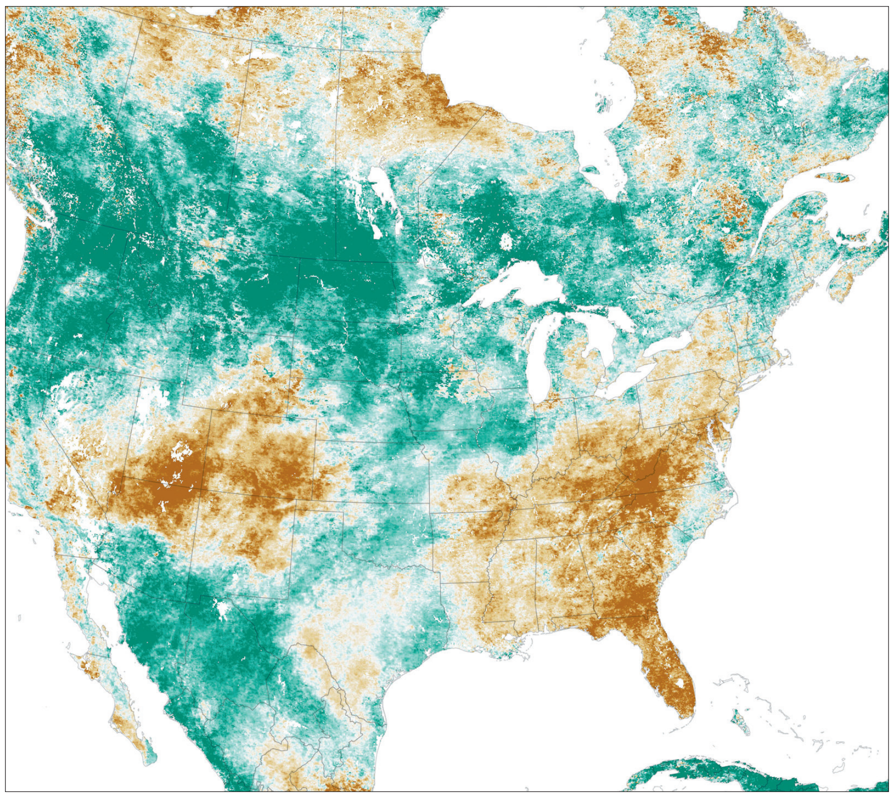
In the coming decades, every major food-growing region faces increasing incidences of sudden water shortage.

Jeffrey Basara and Jordan Christian

Flash droughts develop fast, and when they hit at the wrong time, they can devastate a region's agriculture. They're also becoming increasingly common as the planet warms.

In a study we published this year in the journal *Communications Earth and Environment*, we found that the risk of flash droughts, which can develop in the span of a few weeks, is on pace to rise in every major agriculture region around the world in the coming decades.

In North America and Europe, cropland that had a 32 percent annual chance of a flash drought a few years ago could have as much as a 53 percent annual chance of a flash drought by the final decades of this century. That result would put food production, energy, and water supplies under increasing pressure. The cost of damage will also rise. A flash drought in the Dakotas and Montana in 2017 caused $2.6 billion in agricultural damage in the United States alone.



Over the course of about four weeks in the summer and fall of 2019, the southeast region of the United States went from green to brown, as shown in this NASA Earth Observatory map (*above*) of *evapotranspiration*, a measure of how much water is leaving from the ground surface and vegetation in a region. The area was experiencing what's called a *flash drought*, in which a set of conditions coincide to produce drought very quickly. Unusually warm temperatures, high levels of sunlight, low humidity, windy conditions, and lack of rain can all combine to produce flash droughts, and crops in affected regions can fail if the droughts occur at key times. In 2012, another flash drought stunted the growth of corn in Nebraska (*right*) and in other areas of the central United States. The incidence of flash droughts is only expected to increase because of climate change.

Left: NASA Earth Observatory/Lauren Dauphin/Christopher Hain; right: Pasquale Mingarelli/Alamy Stock Photo

How Flash Droughts Develop

All droughts begin when precipitation stops. What's interesting about flash droughts is how fast they reinforce themselves, with some help from the warming climate.

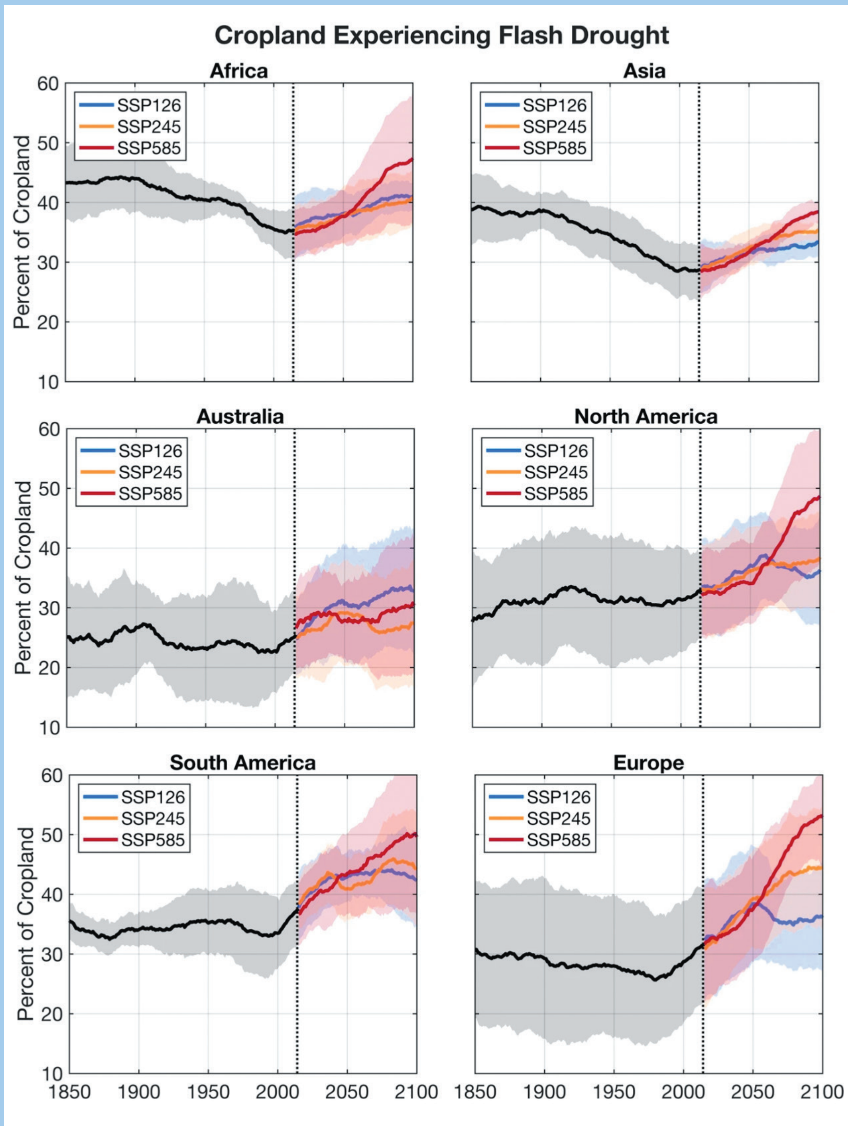
When the weather is hot and dry, soil loses moisture rapidly. Dry air extracts moisture from the land, and

QUICK TAKE

The warming climate has accelerated the already increasing occurrence of flash droughts, which can cause crops to fail as well as put water and energy supplies at risk.

When certain conditions combine, flash droughts can arise over the span of just a few weeks, quickly pulling moisture from the ground surface and plants.

Better prediction of flash droughts could help farmers and ranchers plan for a changing future, which modeling shows will only see further increases in flash droughts.



J. I. Christian et al., 2023

Climate models indicate that more land will be in flash drought in every region of the world in the coming decades. The black line shows historical data. Three different possible future scenarios of greenhouse gas emissions show how low (SSP126), medium (SSP245), and high (SSP585) levels are likely to impact the percent of cropland affected by flash drought. A 30-year centered moving average is applied to each time series; the shaded regions indicate the variability among the averages between all six models for each scenario.

rising temperatures can increase this *evaporative demand*. The lack of rain during a flash drought can further contribute to the feedback processes.

Under these conditions, crops and vegetation begin to die much more quickly than they do during typical long-term droughts.

Global Warming and Flash Droughts

In our study, we used climate models and data from the past 170 years to gauge the drought risks ahead under three scenarios for how quickly the world takes action to slow the pace of global warming.

If greenhouse gas emissions from vehicles, power plants, and other human sources continue at a high rate, we found that cropland in much of North America and Europe would have a 49 and 53 percent annual chance of flash droughts, respectively, by the final decades of this century. Globally, the largest projected increases in flash droughts would be in Europe and the Amazon.

Slowing emissions can reduce the risk significantly, but we found flash droughts would still increase by about 6 percent worldwide under a low-emissions scenario.

Timing Is Everything for Agriculture

We've lived through a number of flash drought events, and they're not pleasant. People suffer. Farmers lose crops. Ranchers may have to sell off cattle. In 2022, a flash drought slowed barge traffic on the Mississippi River, which carries more than 90 percent of the United States' agriculture exports.

If a flash drought occurs at a critical point in the growing season, it could devastate an entire crop.

Corn, for example, is most vulnerable during its flowering phase, which is called *silking*. That stage typically happens in the heat of summer. If a flash drought occurs then, it's likely to have extreme consequences. However, a flash drought closer to harvest can actually help farmers, as they can get their equipment into the fields more easily.

In the southern Great Plains of the United States, winter wheat is at its highest risk during seeding, in September to October the year before the crop's spring harvest. When we looked at flash droughts in that region during that fall seeding period, we found greatly reduced yields the following year.

Looking globally, paddy rice, a staple for more than half the global population, is at risk in northeast China and other parts of Asia. Other crops are at risk in Europe.

Ranches can also be hit hard by flash droughts. During a huge flash drought in 2012 in the central United States, cattle ran out of forage and water became scarcer. If rain doesn't fall during the growing season for natural grasses, cattle don't have food, and ranchers may have little choice but to sell off part of their herds. Again, timing is everything.

It's not just agriculture. Energy and water supplies can be at risk, too. Europe's intense summer drought in 2022 started as a flash drought that became a larger event as a heat wave settled in. Water levels fell so low in some rivers that power plants shut down because they couldn't get water for cooling, compounding the region's problems. Events like those are a window into what countries are already facing and could see more of in the future.

Not every flash drought will be as severe as what the United States and Europe saw in 2012 and 2022, but we're concerned about what kinds of flash droughts may be ahead.

Thierry Monasse/Getty Images News

Flash droughts affect agriculture directly, and they can also interrupt the transport of crops and even people. During Europe's 2022 drought, which began as a flash drought and then intensified, house boats in the Netherlands were left stranded in a dry riverbed.

Can Agriculture Adapt?

One way to help agriculture adapt to the rising risk is to improve forecasts for rainfall and temperature, which can help farmers as they make crucial decisions, such as whether they'll plant or not.

When we talk with farmers and ranchers, they want to know what the weather will look like over the next one to six months. Meteorology is pretty adept at short-term forecasts that look out a couple of weeks, and at longer-term climate forecasts using computer models. But flash droughts evolve in a midrange window of time that is difficult to forecast.

We're tackling the challenge of monitoring and improving the lead time and accuracy of forecasts for flash droughts, as are other scientists. For example, the U.S. Drought Monitor at the University of Nebraska-Lincoln has developed an experimental short-term map that can display developing flash droughts. As scientists learn more about the conditions that cause flash droughts and about their frequency and intensity, forecasts and monitoring tools will improve.

Cropland in much of North America and Europe would have a 49 and 53 percent annual chance of flash droughts, respectively, by the final decades of this century.

Increasing awareness can also help. If short-term forecasts show that an area is not likely to get its usual precipitation, that should immediately set off alarm bells. If forecasters are also seeing the potential for increased temperatures, that heightens the risk for a flash drought's developing.

Nothing is getting easier for farmers and ranchers around the world as global temperatures continue to rise. Understanding the risk they could face from flash droughts will help them—and anyone concerned with water resources—manage yet another challenge of the future.

Bibliography

Christian, J. I., et al. 2023. Global projections of flash drought show increased risk in a warming climate. *Communications Earth and Environment* 4:165.

Christian, J. I., et al. 2021. Global distribution, trends, and drivers of flash drought occurrence. *Nature Communications* 12:6330.

Otkin, J. A., et al. 2018. Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bulletin of the American Meteorological Society* 99:911–919.

Jeffrey Basara is an associate professor at the University of Oklahoma with a joint appointment between the School of Meteorology and the School of Civil Engineering and Environmental Science. Jordan Christian is a postdoctoral researcher in meteorology at the University of Oklahoma. This article is adapted and expanded from one that appeared in The Conversation, *theconversation.com. Email for Basara: jbasara@ou.edu*

On the Hunt for Another Earth

Astronomers are making progress in finding planets that broadly resemble our own.

Abel Méndez

Extraterrestrial life is a genuine scientific possibility. Earth proves that at least one habitable planet, even a planet with technological intelligence, can exist. Astronomers have discovered more than 5,000 exoplanets—worlds circling other stars—and that number increases weekly. Now the search for habitable worlds is on, pushing our data and models to the limits.

Most known exoplanets have been detected indirectly, primarily using two techniques. The *transit method* measures the slight dimming of a star that occurs when a planet passes in front of it. The *radial velocity method* measures tiny shifts in a star's light spectrum caused by the back-and-forth gravitational pull of orbiting planets. The amount of light blocked during a transit indicates the size of a planet. The intensity of a radial velocity change indicates a planet's approximate mass. When combining the two methods is possible, the results reveal the planet's density. And that is where our knowledge of most exoplanets ends.

To complement the search for Earth-like planets, I have created an Earth Similarity Index (ESI) that indicates how much planets resemble Earth in their gross physical properties. The ESI is a number between zero (no similarity) and one (identical to Earth). It is not a direct measure of habitability, but rather a fuzzy comparison between a selected set of planetary properties with Earth—primarily size, mass, and *insolation*, the amount of radiation received from the planet's star—because those details are often all that we are able to measure for exoplanets. Worlds with high ESI values are not necessarily more habitable, because so many other factors influence a planetary environment. Venus (ESI = 0.78) and Mars (ESI = 0.65) are similar in size and insolation to Earth, but they have starkly different surface conditions.

Most known exoplanets have ESI values well below that of our Moon (ESI = 0.60). There are a few exoplanets with ESI values close to one. Maybe their surfaces are genuinely similar to Earth's, with oceans and continents. Then again, they could be unsuitable for any life because they lack other life requirements, such as water, or have harmful or toxic conditions. We don't know how likely it is that superficially Earth-like planets truly resemble our own; that is a major goal for the next few decades of astronomical observations. The ESI is also relevant only for surface life, not subsurface life that could exist independent of the star's energy, much like the life around geothermal vents in Earth's deep oceans. For example, astrobiologists consider it possible that there could be life in the oceans under the icy surface of Jupiter's moon Europa, even though it has an ESI of 0.22.

We are still very early in the process of identifying exoplanets that are not just broadly Earth-like but that seem plausibly habitable. The first claimed potentially habitable exoplanets, discovered around the nearby red dwarf star Gliese 581, turned out to be illusions. Then in 2012, a team of astronomers led by Guillem Anglada-Escudé of the University of Göttingen in Germany announced the discovery of Gliese 667C c, the first confirmed potentially habitable world. Today, we know of about 60 potentially habitable planets that have the right size and orbit to hold surface water. However, the world "potentially" carries a lot of weight here. We do not know anything about these planets' atmospheres or surfaces yet. (*See "Decoding Light from Distant Worlds," May–June 2020.*) They could be barren, Moon-like worlds, or something else unexpected.

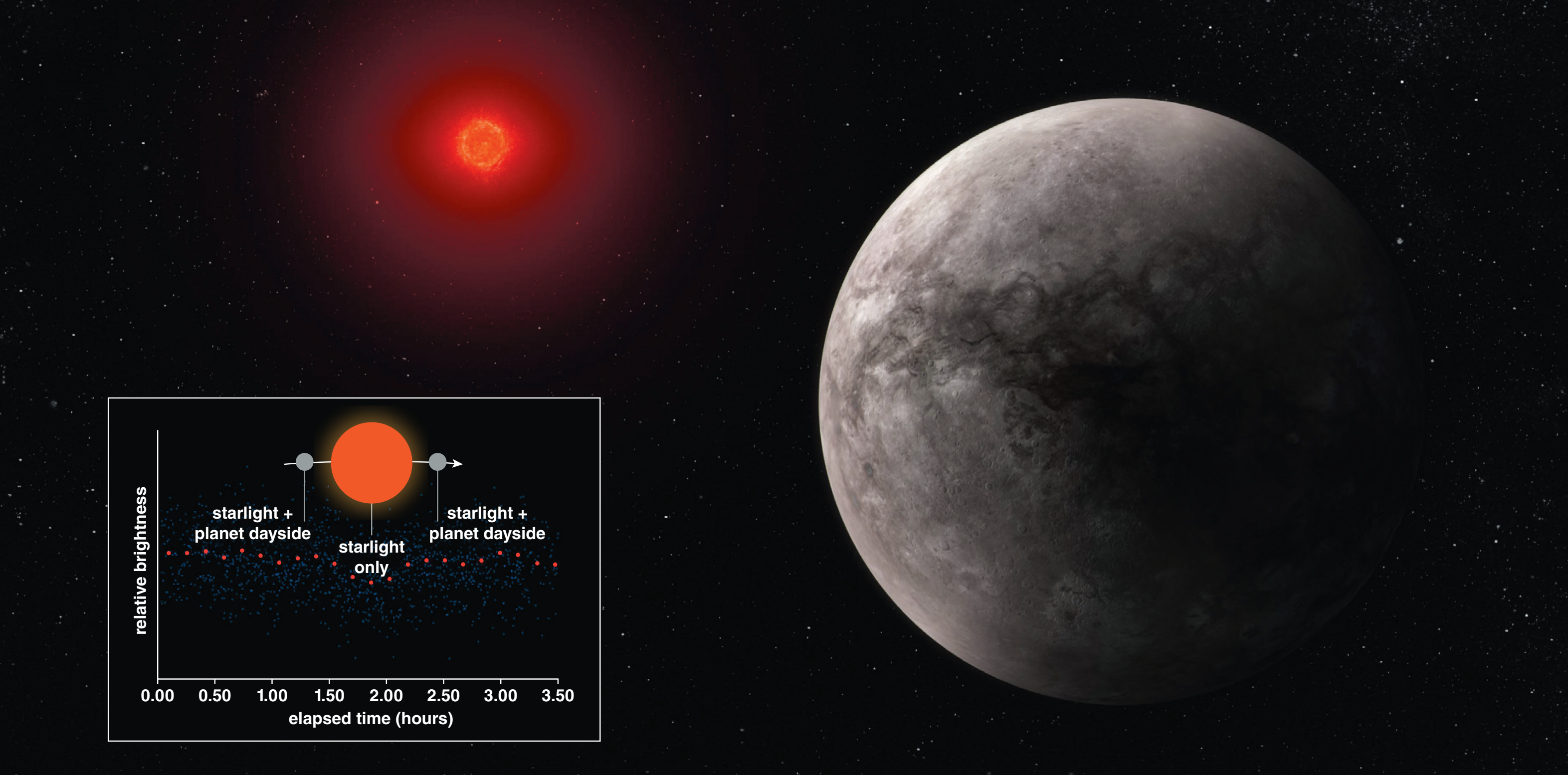
Even today's most powerful telescopes, such as the James Webb Space Telescope (JWST), are able to analyze the atmospheric composition of only a few, particularly well-situated exoplanets. We will need far better telescopes, such as NASA's proposed Habitable Worlds Observatory, to discover more Earth-like planets, observe some of them directly, and provide more information about their atmospheres, surfaces, and suitability for life. Detailed theoretical studies about habitability and planetary modeling will then be es-

QUICK TAKE

Potentially habitable planets require rocky surfaces with liquid water and a substantial but not overwhelming atmosphere—at least, those are requirements for life as we know it.

No true Earth analog has yet been found, but the author has developed an Earth Similarity Index to identify planets that broadly resemble our own and deserve closer scrutiny.

Gases associated with life on Earth, such as oxygen alongside methane, could indicate biological activity on another planet. The search for chemical *biosignatures* is starting now.



NASA, ESA, CSA, J. Olmsted. Data: Thomas Greene, Taylor Bell, Elsa Ducrot, Pierre-Olivier Lagage

Planets around red dwarf stars are relatively easy to study, making them natural starting points in the search for habitable conditions. Recent observations of planet TRAPPIST-1 b (*illustrated here*) detected no atmosphere when it passed in front of its star (*inset*), but models had predicted it would be airless. Future studies will target TRAPPIST-1's six other Earth-sized planets, three of which orbit in the habitable zone around their stars.

sential for interpreting the results. The confident detection of life won't be easy.

Using current technology, we would have to get extremely lucky to get a clear indication of life, or even a clear signal of habitability. Then again, if we *do* get lucky, we might start to get an answer as soon as this decade. Even if it takes us a long time to get there, we are on the path to discovery—which is truly exciting.

What Life Needs

In astrobiology, habitability is defined as the overall suitability of an environment for biological processes. A habitability analysis cannot determine whether a world is able to sustain life; only life detection experiments can do that. Even with that restriction, researchers cannot study all the factors affecting habitability, so we focus on the ones that are most important and that are relatively straightforward to measure.

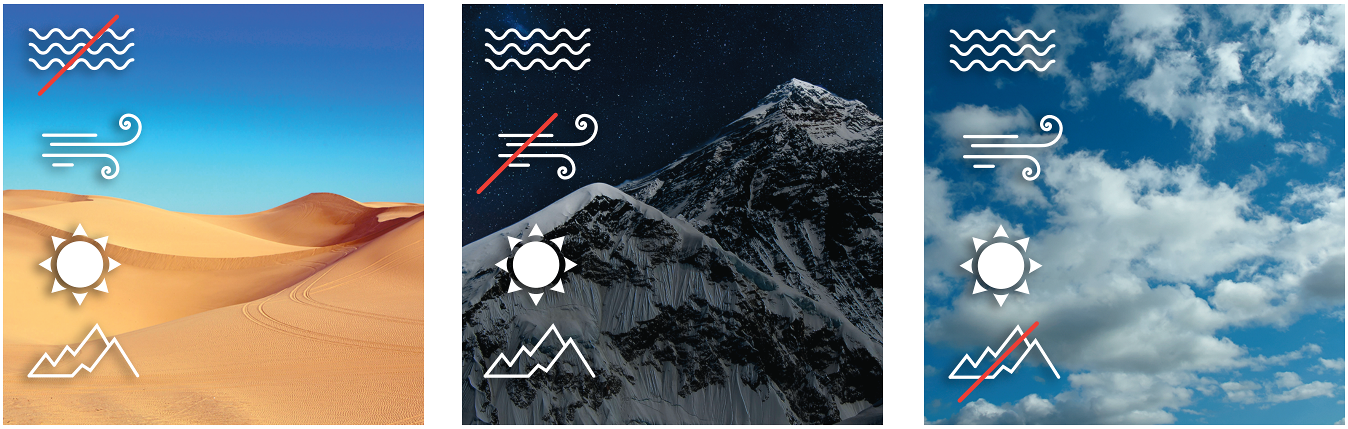
The right temperature is a key consideration for habitability. Most animal life on Earth requires temperatures between 0 degrees to 50 degrees Celsius, and the optimum is around 25 degrees, not far from Earth's average global temperature of 15 degrees. Microbial life has an extended thermal range if the water is still liquid. We don't know whether this principle of biochemistry is universally applicable to all extraterrestrial life, or if it is a result of the evolution of our particular planet. Nevertheless, liquid water is essential to life on Earth, so astrobiologists start from the assumption that habitable exoplanets must have a surface temperature compatible with liquid water.

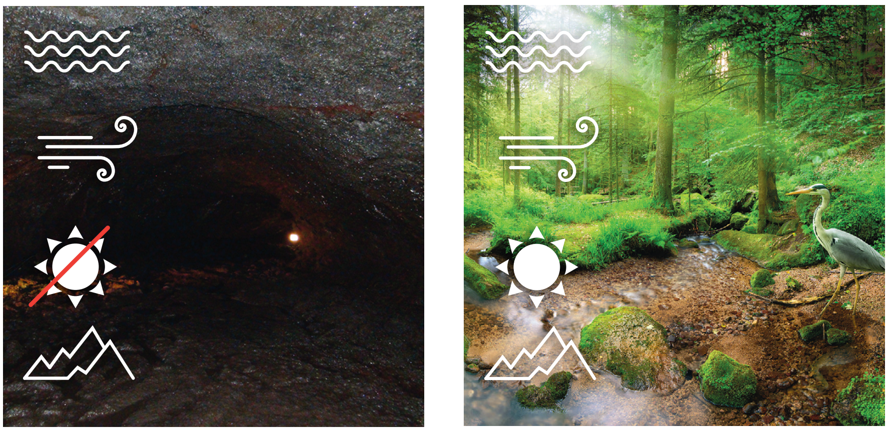
Life has many requirements beyond water, however. It requires an environment where the three phases of matter (gas, liquid, and solid) plus energy coexist for a long time. The classical elements (air, water, earth, and fire) are helpful analogies for understanding this requirement.

The three phases of matter are necessary because the elements used to construct the molecule of life, DNA (hydrogen, carbon, nitrogen, oxygen, and phosphorus), are not present in a single phase. For example, nitrogen is usually found as a gas in the atmosphere, whereas inorganic sulfur and phosphorus are typically found in rocks as sulfates and phosphates. Furthermore, the solvent, diffusion, and flow properties of air and water enable these elements to mix with rock and become readily available in any habitat. On a planetary scale, the classical elements become the atmosphere (air), hydrosphere (water), lithosphere (earth), and insolation (fire).

Exoplanet surveys indicate that there are, on average, one to two planets per star—and yes, that includes all the stars we see at night. What fraction of those stars have a habitable world is an open question. The rarer such planets are, the less likely we are to have one close enough to study using current telescopes. The term *Eta-Earth* is used to refer to the average number of Earth-sized planets, in orbits that allow Earth-like temperatures, around each star in our galaxy. NASA's Kepler space telescope, which operated from 2009 to 2018, was the first mission to provide empirical estimates of Eta-Earth. Upper estimates of Eta-Earth put the value around 0.5, which means that less than half of the stars we see at night might have planets with the fundamental properties for habitability: not just a suitable orbit, but also a suitable size.

The right size is another key habitability factor, because life as we know it needs a rocky surface and a substantial but not excessive atmosphere. Planets in the habitable zone must be between 0.5 (Mars-sized) and 1.5 times the diameter of Earth to meet those requirements. Worlds that are too small will have a rocky surface but won't be able to hold a significant atmosphere, just like Mars. Worlds that are too large will have too much atmosphere and no accessible rocky surface; an atmosphere of just 1 percent of the planet's mass would create pressures so extreme that surface water would remain solid no matter what the temperature.





The requirements of life mirror the four elements of classical philosophy: air, water, earth, and fire. Without water (*upper left*), known biochemistry is impossible. Without air (*upper middle*), there are no resources for respiration. Without a rocky surface (*upper right*), life lacks crucial minerals. Without the light of fire (*far left*), life lacks a primary energy source. The combination of all four elements is what makes Earth so vibrant (*left*). That's why forests are the most hospitable environments on Earth. Habitable planets should also have forests to support a large diversity of animals, including intelligent life.

Abel Méndez/adapted by Barbara Aulicino; photos: Wikimedia Commons (3), Travel Stock/Shutterstock, Smileus Images/Alamy Stock Photo

The right orbit of a planet around its star is related to both the size and temperature requirements for habitability. The *stellar habitable zone* is defined as the region around a star where orbiting planets potentially receive the right amount of warming radiation to support liquid water on their surfaces. This region is sometimes called the "Goldilocks zone" because planets in that zone are neither too hot nor cold, but are just right for life as we know it. If a planet is too close to its star, its surface temperature will be too high for liquid water to exist, whereas if it is too far away, the temperatures will be too low. (Those considerations don't exclude subsurface life, however.)

The boundaries of the habitable zone depend on the effective temperature and luminosity of the star, but also on the planet's properties, such as its size and its atmospheric composition. For example, the boundaries for a star like our Sun extend from about 0.95 to 1.37 astronomical units (the distance between Earth and the Sun), but the zone could be smaller or larger depending on the specific characteristics of a planet. Planets with thick, heat-trapping atmospheres, like Venus, may be able to support liquid water at greater distances from their stars than planets with thin atmospheres, like Mars.

Surveys indicate that there are, on average, one to two planets per star—and yes, that includes all the stars we see at night.

The inner edge of the habitable zone is constrained by warming induced by the evaporation of water, a strong greenhouse gas. Once the average temperatures exceed about 60 degrees, a runaway feedback effect causes all the oceans to be converted into steam. Stellar wind and radiation will then erode this atmospheric water vapor, decomposing it into hydrogen (which easily escapes into space) and oxygen (which quickly reacts with other atmospheric or surface compounds). Venus may have had early oceans that were lost in this way. The outer limit of the habitable zone is constrained by the greenhouse capacity of carbon dioxide to keep temperatures above freezing. Mars, with its relatively low gravity, lost its water and most of its initial atmosphere, leaving it cold and inhospitable.

The right type of star is necessary to have a long-lived, potentially Earthlike planet. Small stars generally have small planets, whereas big stars have more giant planets. About 75 percent of the stars in our galaxy are red dwarf stars (known to astronomers as type M stars), which are more prone to having Earth-sized planets. Only 19 percent of stars are Sunlike (referred to as types F, G, or K); they are more likely to hold a mix of Earth-sized planets and giant worlds like Jupiter and Saturn. Our Sun is a typical G star, but we don't know if G stars are uniquely or even especially well-suited to life.

There are different habitability issues associated with Sunlike stars and with red dwarf stars. Sunlike stars are more likely to have giant planets that might precipitate early rains of life-disrupting asteroids, or even eject smaller planets from their orbits during their formation. Red dwarf stars, despite being smaller and dimmer, are more active than Sunlike stars, especially during their early stages; they produce powerful flares that could eventually erode the atmosphere of their planets. Because red dwarfs are dim and cool, any potentially habitable planets would need to orbit very close to their star. In such orbits, planets are likely to become tidally locked, meaning that one hemisphere always faces the star while the other hemisphere always faces away, trapped in eternal darkness. On such planets, it's unclear if life could survive in the twilight zone between the hemispheres, or how likely it is that their atmosphere and oceans could even out the extreme day–night temperature differences.

Habitable worlds also depend on many other stellar and planetary properties. For example, the stars need to remain stable long enough for life to start and evolve on their planets. F stars have lifetimes that are somewhat shorter than our Sun's. One billion years might be enough for microbial life to develop. Animals might require a few billion years more. Earth is 4.6 billion years old, and we know of exoplanets younger and older than this. There is no guarantee that older planets should be better for life. Stars tend to get brighter with age, which could move a planet out of the habitability zone. Astronomers estimate that Earth will be too hot for oceans in less than 2 billion years. K stars have longer stable lifetimes than the Sun, and M stars have the longest of all—in some cases, greater than a trillion years.

Planets also need to have the right composition and internal structure. On Earth, oceans, continents, and plate tectonics help stabilize the thermal environment and recycle nutrients. A global magnetic field protects our atmosphere from erosion by the solar wind, and the Moon helps stabilize Earth's axis tilt and climate. All these factors might not be essential to support life, but they seem to help prolong it, allowing time for the evolution of more complex forms.

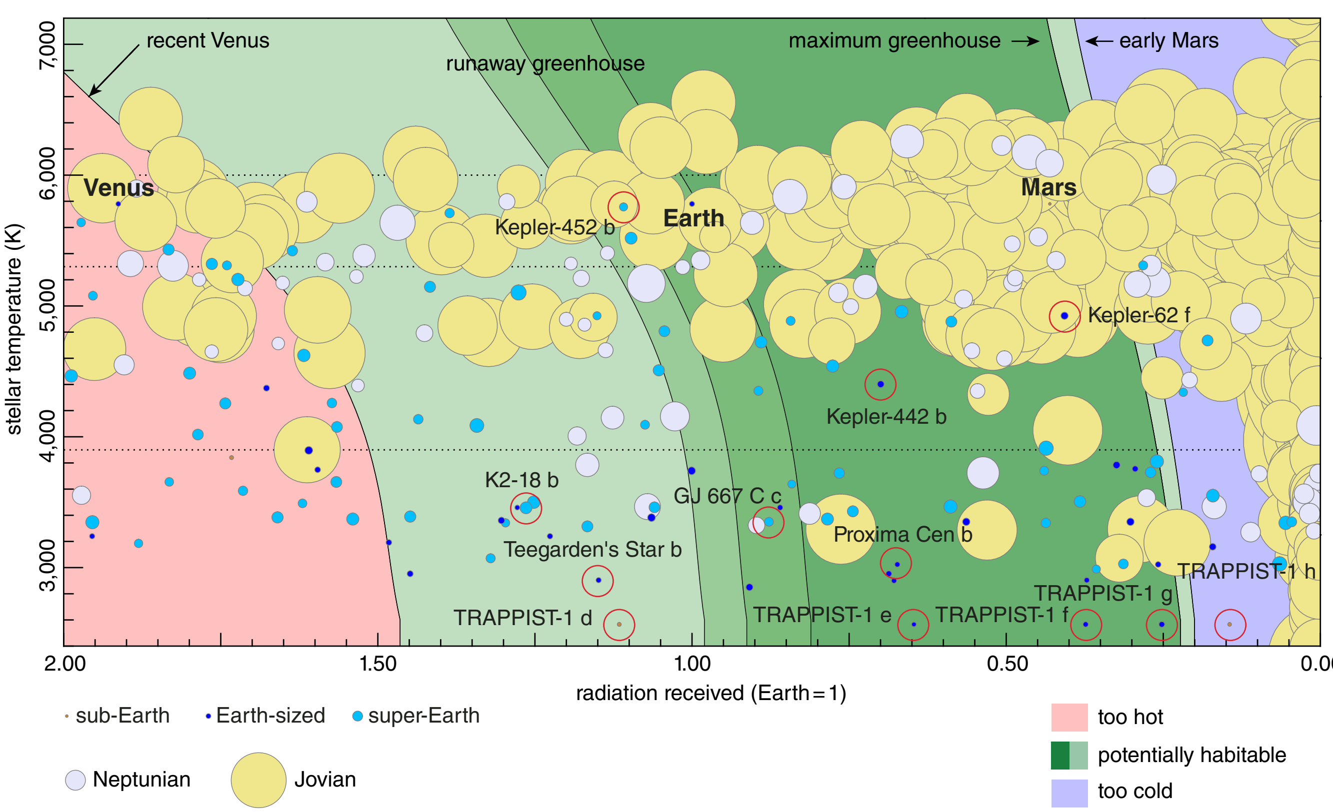
Targeting Habitable Worlds

Since 2011, my colleagues and I at the Planetary Habitability Laboratory of the University of Puerto Rico at Arecibo have maintained the Habitable Exoplanets Catalog to track the number of potentially habitable worlds discovered by space-based and ground-based telescopes. The criteria for inclusion in the catalog are that the exoplanet is approximately Earth-sized and orbiting within its stellar habitable zone. We use the ESI to sort them out by their resemblance to Earth as determined by their diameters and by the amount of radiation they receive from their star. We don't consider the possibility of subsurface life, because it would be almost impossible to detect.

In practice, exoplanets with ESI values above 0.5 meet our broad criteria for habitability. The Habitable Exoplanets Catalog also includes estimates of the planets' possible surface temperatures, assuming that they are rocky and have a terrestrial atmosphere. If

A plot of the sizes and *insolations* (amount of stellar radiation received) of known planets indicates the wide range of other worlds out there. Green marks the stellar habitable zone, where a planet could be the right temperature to have liquid water; darker green indicates higher confidence in habitability. Being within the habitable zone does not guarantee life-friendly conditions, however. Mars orbits within the zone, for instance, but it is so small that it has lost most of its atmosphere and turned cold and dry.

Adapted from Abel Méndez



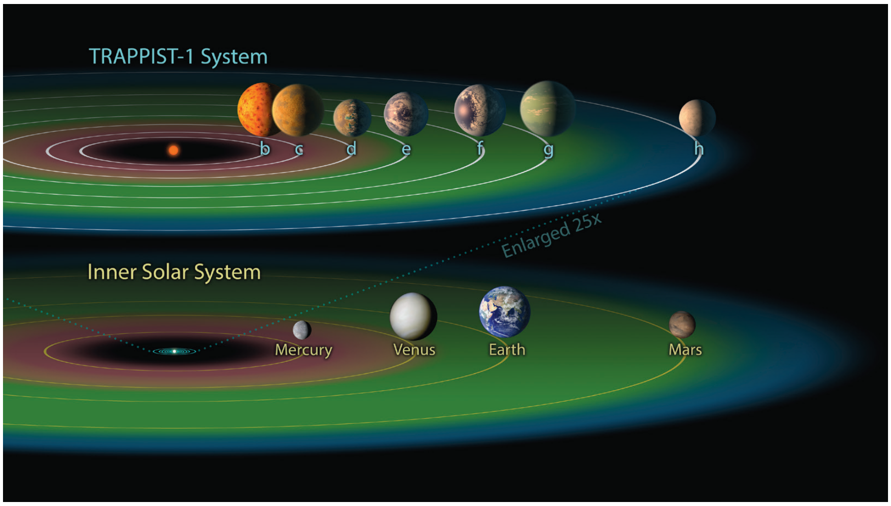


Illustration: NASA/JPL-Caltech

The TRAPPIST-1 system (*top*) contains seven Earth-sized planets circling a red dwarf star. Although at least three of these planets could have habitable temperatures, they are surely quite unlike anything in our solar system (*bottom*). TRAPPIST-1 is a dim, cool red star, and its planets huddle tightly around it. Such stars are extremely common, but we don't know yet whether their planets are capable of supporting life.

the planets have no atmosphere, they would be colder than those estimates; if they have denser atmospheres, they would be hotter. The big question is how many of these worlds are indeed Earth-like, if any.

At present there are up to 63 potentially habitable exoplanets on our list, out of the thousands that have been discovered. Keep in mind that for these planets, we know only their orbits, along with their sizes or masses. Those details suggest they could support surface liquid water, but planets can lose their atmosphere and water during their history and become desert worlds. For example, our Moon is next to Earth and in the Sun's habitable zone, but it is too small to hold any atmosphere. Other planets could be ocean worlds that have too much water for life because they are diluting the availability of the other crucial ingredients. Planets can also have too much atmosphere, preventing the formation of liquid water.

The current list of 63 potentially habitable exoplanets might include deserts, oceans, gas planets, or some with land and oceans, just like Earth. All we know is that these planets have at least the basic potential for solid land and surface oceans in which the three phases of matter are in contact, with plenty of energy from their star. The Habitable Exoplanets Catalog includes planets up to 2.5 times the diameter of Earth to be more inclusive of ocean worlds and to allow for errors associated with radius estimates. About 24 exoplanets on our list are more likely to be of a rocky composition, because they have a diameter less than 1.5 times that of Earth. Something interesting is going on with these planets, which is what astronomers are trying to figure out.

A nearby star system called TRAPPIST-1 is the best natural lab we've found for putting our models of planetary habitability to the test. At the center is a tiny red star, 40 light-years from Earth in the constellation Aquarius, surrounded by seven Earth-sized planets (a light-year is the distance that light travels in 1 year, 9.46 trillion kilometers). This system is named after the Transiting Planets and Planetesimals Small Telescope (TRAPPIST) in Chile, which first detected the planets in 2016. TRAPPIST-1 is an ultracool dwarf star with a size and mass about one-tenth that of the Sun. Three of its planets—designated "b," "c," and "d"—orbit in the inner part of the habitable zone and are likely too hot to support life, whereas planet "h" is probably too cold. But planets "e," "f," and "g" orbit well within the habitable zone. (Someday, these exoplanets will have proper names, as the International Astronomical Union is working on exoplanet naming with the public's help.)

TRAPPIST-1 conveniently has hot, temperate, and cold Earth-sized planets in one system. Better yet, all seven planets transit in front of their parent star, so we can observe their shadows, and the system is close enough to Earth that we can study their atmospheres using JWST. By looking at just this one star long enough, we could obtain some extremely revealing information about the atmospheric diversity of Earth-sized exoplanets. That said, we should be clear that the TRAPPIST-1 system is very different from our own. It is a great test case for planets around the abundant red dwarfs, but it won't tell us much about true Solar System analogs.

Astronomers study the atmospheres of exoplanets using a technique called *transit spectroscopy*. When a planet transits its star, some of the star's light passes through the planet's atmosphere before reaching us. We can learn about the chemical composition and physical properties of the planet's atmosphere by analyzing the light that has passed through the atmosphere. This technique requires making extremely accurate measurements of the star's spectrum before, during, and after the transit. Comparing the combined spectrum with the spectrum of the star alone can reveal any differences caused by the planet's atmosphere. These differences can then be linked to known molecules, such as water vapor, methane, and carbon dioxide.

Transit spectroscopy has been successfully used to study a wide range of exoplanets. Most successful studies so far have been done on giant planets, but the situation is changing. In 2019, the Hubble Space Telescope detected, for the first time, water vapor in the atmosphere of a small planet orbiting in a habitable zone. This discovery, on exoplanet K2-18 b, sparked interest in it as a target for further observations to search for signs of life. However, K2-18 b has a diameter 2.3 times that of Earth and is eight times as massive, suggesting that it is something between a water world and a mini-Neptune, not a rocky planet.

Water is essential for life as we know it, but it does not necessarily suggest the presence of life; water vapor is ubiquitous in planetary atmospheres. Oxygen and methane are more revealing atmospheric *biosignatures*—chemicals that could indicate the presence of significant, global biological activity on another planet.

Oxygen is a highly reactive gas that, on Earth, is produced by photosynthesis in plants and by some bacteria. No other planet in the Solar System has significant quantities of free oxygen in its atmosphere. If we detect oxygen in the atmosphere of an exoplanet, that could be an indicator of the presence of photosynthetic life. But atmospher-

ic oxygen could also be produced by nonbiological processes, such as the photolysis of water or the breakdown of ozone. Detecting oxygen alone is therefore not enough to confirm the presence of life on another world.

Methane gas can also be produced by both living and nonliving processes. On Earth, most of the methane is produced biologically. It is released in great quantities by certain microbes, including methanogenic archaea, as part of their metabolic processes. But it can also arise abiotically through processes such as *serpentinization* (a reaction between minerals and water), the breakdown of organic matter, and volcanic activity.

Things would get truly interesting if we were to find methane and oxygen together. Methane breaks down quickly in the presence of oxygen, so finding both gases together would indicate that some process keeps replenishing both of them. Biology is the most effective process we know of that is able to do that. If we detect significant quantities of oxygen and methane in the atmosphere of another planet, it could be a strong indicator of the presence of life.

top 10 potentially habitable exoplanets

	name	mass (Earths)	temperature (Kelvin)	period (days)	distance (light-years)	Earth similarity
1	Teegarden's Star b	≥ 1.05	~ 298	4.9	12	0.95
2	TOI-700 d	~ 1.57	~ 278	37.4	101	0.93
3	Kepler-1649 c	~ 1.20	~ 303	19.5	301	0.92
4	TRAPPIST-1 d	0.39	~ 296	4.0	41	0.91
5	LP 890-9 c	—	~ 281	8.5	106	0.89
6	Proxima Centauri b	≥ 1.27	~ 257	11.2	4.2	0.87
7	K2-72 e	~ 2.21	~ 307	24.2	217	0.87
8	GJ 1002 b	≥ 1.08	~ 261	10.3	16	0.86
9	GJ 1061 d	≥ 1.64	~ 247	13.0	12	0.86
10	GJ 1061 c	≥ 1.74	~ 311	6.7	12	0.86

PHL @ UPR Arecibo

The most Earth-like planets discovered so far are ranked according to their Earth Similarity Index, a broad measure of their sizes, masses, and the amount of warming radiation they receive. All 10 planets orbit red dwarf stars. Several of these planets transit in front of their stars, creating a favorable geometry for atmospheric analysis. The TRAPPIST-1 planets are especially interesting because they transit, are nearby, and three of them have a high Earth Similarity Index.

The Search Begins

JWST should be able to search for signs of water, methane, and carbon dioxide in the atmospheres of the planets around TRAPPIST-1. It will examine a handful of other Earth-sized exoplanets as well. Detecting oxygen lies right at the edge of the telescope's capabilities, but it is possible—if we get lucky—that JWST will find evidence of both oxygen and methane on an exoplanet. That finding would be the first significant evidence of life beyond our Solar System.

Such a discovery would surely be controversial, and hard to verify independently. An unambiguous detection of oxygen will probably require much bigger telescopes, either on the ground or in space. The challenge is even greater for some other high-interest biosignature gases such as phosphine, which is produced by anaerobic microbes on Earth.

For now, my colleagues and I are approaching the problem step by step. Because red dwarf stars are highly active and could potentially erode the atmospheres of their planets, it would be interesting enough to find *any* atmosphere in the TRAPPIST-1 planets. Recent transit spectroscopy observations by JWST did not detect an atmosphere around planet "b," but that was expected because this planet is too close to its star. Future observations of other planets in the TRAPPIST-1 system will greatly clarify the prospects for life on worlds around red dwarf stars.

Astronomers are also working on different techniques for studying the atmospheres of potentially habitable exoplanets. Currently, we can get atmospheric data only for exoplanets that transit their host star. The edge-on alignment necessary for transits to occur happens in less than 1 percent of planetary systems, so we are missing a lot. The planned Habitable Worlds Observatory, along with next-generation ground-based telescopes such as the Extremely Large Telescope under construction in Chile, will overcome these limitations. These telescopes will be able to take direct images of exoplanets in the habitable zone, incorporating an occulter or *coronagraph* to block the vastly brighter light of the parent star.

Direct imaging using a variety of techniques will dramatically increase the number of exoplanets we can target, including planets around Sunlike stars and the Earth-sized planet orbiting our nearest stellar neighbor, Proxima Centauri. By observing potentially habitable exoplanets directly, we can also expect a wealth of new information about their atmospheres and surface properties, such as temperatures and the presence of oceans. Those insights, in turn, will deepen our understanding of habitability and help us make sense of the presence, or absence, of possible biosignatures. At the Planetary Habitability Laboratory, we are developing models to determine what combination of atmospheres, temperatures, and ocean-to-land fractions is most suitable for life. We assume that planets with high ESI values are more likely to be habitable, but we may learn that is not always true.

Whatever we find, the results will be transformative. In the coming decades, we will be able to say, for the first time, how unusual Earth and its life are in the galaxy.

References

Billings, L. 2013. *Five Billion Years of Solitude: The Search for Life among the Stars*. New York: Penguin.

Méndez, A., et al. 2021. Habitability models for astrobiology. *Astrobiology* 21:1017–1027.

Méndez, A. 2022. Habitable exoplanets. In *New Frontiers in Astrobiology*, eds. R. Thombre and P. Vaishampayan, pp. 179–192. Cambridge, MA: Elsevier.

Schulze-Makuch, D., et al. 2011. A two-tiered approach to assessing the habitability of exoplanets. *Astrobiology* 11:1041–1052.

Thompson, M. A., J. Krissansen-Totton, N. Wogan, M. Telus, and J. J. Fortney. 2022. The case and context for atmospheric methane as an exoplanet biosignature. *Proceedings of the National Academy of Sciences of the U.S.A.* 119:e2117933119.

Abel Méndez is the director of the Planetary Habitability Laboratory and professor of physics and astrobiology at the University of Puerto Rico at Arecibo. He uses astronomical observations, theoretical models, and computer simulations to study the habitability of Earth, other worlds of the Solar System, and exoplanets. Email: abel.mendez@upr.edu

SCIENTISTS' Nightstand

The Scientists' Nightstand, American Scientist*'s books section, offers reviews, review essays, brief excerpts, and more. For additional books coverage, please see our* Science Culture *blog channel online, which explores how science intersects with other areas of knowledge, entertainment, and society.*

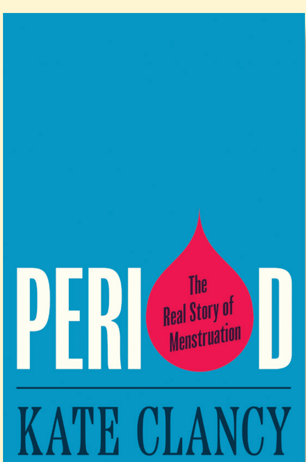
ALSO IN THIS ISSUE

MODELS AND MATHEMATICS: Q&A WITH ERICA THOMPSON
page 251

ONLINE

On our *Science Culture* blog:
www.americanscientist.org/blogs/science-culture

Rethinking Menstrual Norms
Digital features editor Katie Burke reviews Kate Clancy's *Period: The Real Story of Menstruation*.



Glitching Out

Ashley Shew

More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech. Meredith Broussard. 248 pp. The MIT Press, 2023. $26.95.

The biases caused by and perpetuated through algorithms are no glitch: They are consistent with our world as it exists now. In *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*, Meredith Broussard offers a thorough exploration of how algorithmic systems encode and perpetuate injustice—and what we can do about it.

Broussard's 2018 *Artificial Unintelligence: How Computers Misunderstand the World* demonstrated how we've been misled by tech enthusiasts and developers in their quest for artificial general intelligence (AGI). *More than a Glitch* is no less important to understanding where we are now with tech. Explaining the current state of algorithmic injustice, she offers strong calls to action, along with very concrete steps, to help us manage, mitigate, and recognize the biases built into algorithms of all types. A thread running through the book is the idea of *technochauvinism*, a "bias that considers computational solutions to be superior to all other kinds of solutions." This bias insists that computers are neutral and that algorithms are fair. But this bias is its own type of ignorance, because what we build into algorithms captures and perpetuates existing social biases, all while being touted as "fairer" because of the assumed neutrality of algorithms.

Broussard defines glitch as "a temporary blip," which is how tech boosters often want to explain away publicized instances of bias in tech as one-offs that can be dealt with through individual fixes. For instance, the National Football League (NFL) famously used race in its algorithms to determine payouts for players with brain injuries, with Black players receiving lower amounts by thousands of dollars. The NFL's individual fixes included eliminating data such as race, and paying penalties, although only after they were ordered to by a federal judge. But algorithmic biases are more than a temporary blip or something we can deal with through individual patches; simply fixing blip by blip is not a structural solution and won't let us move quickly enough. Our biases are systemic and structural, and they are reflections of the world.

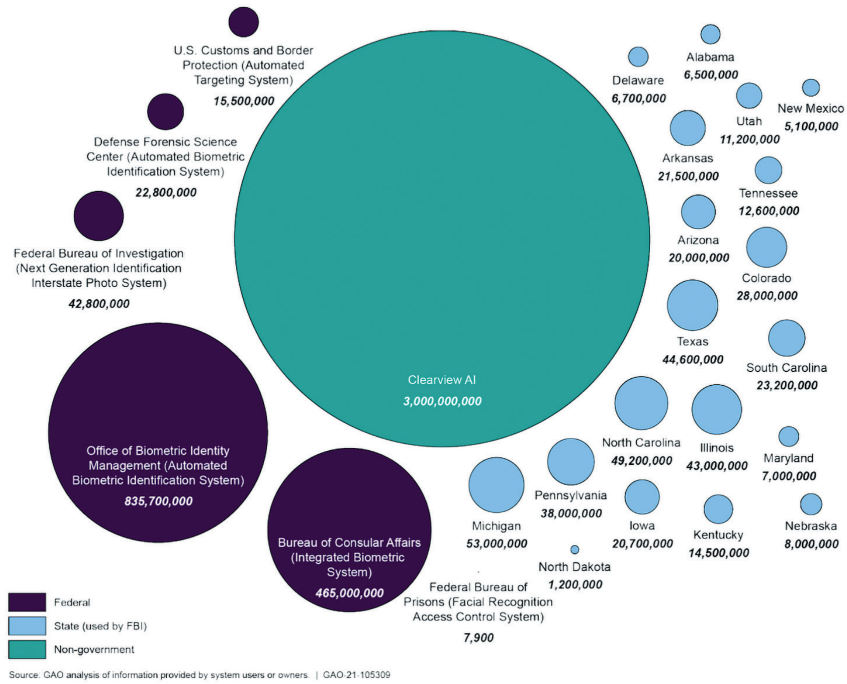
By the end of the book, Broussard offers clear ways to recognize and manage biases and work toward a more just world through public interest technology and the use of algorithmic audits. She highlights the work of groups already doing tech monitoring and reporting, along with many recent cases in which tech companies have failed us, and she suggests how we can add to these efforts through different types of algorithmic auditing and an emphasis on public interest technology. She calls for the public to "hold algorithms and their creators accountable" to the public, and demonstrates ways we can do just that.

The book is divided into 11 chapters, each of which could be read on its own or in concert with other readings. Each chapter is grounded in many contemporary cases of biases built into algorithms that then perpetuate or cause discriminatory harm. Her journalistic approach makes this work come alive through accounts of not only the way algorithms work, but also of how the use of algorithms has caused real harm, including arrests, educational setbacks, administrative nightmares, and even death.

One compelling example from Chapter 3 is from the area of facial rec-

ognition and its use in policing. Broussard explains that facial recognition is far less effective than most people imagine, but police precincts continue to invest money in this technology and try to justify its use. And people's lives are at stake. Broussard shares the story of Robert Williams, who was called into a precinct by the Detroit police, which he thought was a crank call. They arrested him later in the day at his home, flashing an arrest warrant, to the confusion of Williams and his family. He was held for 18 hours without anyone telling him the details of why he had been arrested.

It turned out that Williams had been "identified" as having stolen $3,800 in watches from a grainy image fed into DataWorks Plus software sold to Michigan State Police, but the thief wasn't him. The database used by DataWorks is huge, consisting of 8 million criminal photos and 32 million U.S. Department of Motor Vehicle (DMV) photos. The grainy image from the robbery of someone in a baseball cap was matched by the algorithm with Robert Williams's DMV photo and taken as truth by police: A clear example of unchecked technochauvinism.



This graphic from *More than a Glitch* shows systems with facial recognition technology that are used by federal, state, and nongovernment agencies that employ law enforcement officers, and the number of photos stored in these systems. (Figure from Government Accountability Office report GAO-21-105309.)

Chapter 4 details the case of Robert McDaniel, a Chicago resident who was identified by predictive policing software as someone likely to be involved in a shooting (here, individual police departments aren't specific about what software they use, but CompStat is one example; software giants such as Palantir, Clearview AI, and PredPol make predictive policing software products that are marketed to large police departments). The algorithm wasn't specific about whether McDaniel was supposed to be a shooter or the victim, but the result of identifying McDaniel provoked police visits where they "wanted to talk" to him, social workers being brought in by police, and the police paying close attention to him. This attention was unwanted and dangerous, as he explained to police. He was shot two different times because people began to think he was a snitch or informant, due to the police attention. McDaniel continued to explain that they were putting him in danger, but the Chicago police saw their software as validated. Broussard explains that we've spent "taxpayer dollars on biased software that ends up making citizens' lives worse" due to the belief that such technologies will make things safer and fairer (very often, they won't). For the technologies offered in Chapters 3 and 4, the call to action is to move away from this type of technology development altogether.

But the book is about much more than policing. Chapter 5 explores algorithmic and predictive grading in schools. Algorithms, Broussard argues, can also fail at making predictions that are ethical or fair in education. She presents a number of cases of students harmed through the use of grading algorithms that factor in location and subsequently award lower grades to students in lower-income areas. During the early part of the COVID-19 pandemic, many aspects of life began happening online, including taking standardized tests such as AP, SAT, ACT, and IB exams. In the case of IB (International Baccalaureate) exams, in-person testing was not possible and an algorithm was created to give students grades based on algorithmic predictions. They used a standard practice in data science: Their solution was to take all available data on students and their schools, and to feed it into their system and construct a model to output individual grade predictions. Broussard shares the case of Isabel Castañeda, a student hoping to earn credit for a high IB score in AP Spanish. Castañeda is fluent in English and Spanish, and has studied Korean for years. Her IB Spanish grade was a failing grade. Essentially, she was punished for where she lived and went to school, a school with very low success rates on IB exams. The use of such programs have real impacts in terms of individual costs, such as having to retake tests and college classes, as well as the colleges to which one is accepted. Broussard urges communities and administrators to "resist the siren call" of these systems, however appealing they may be.

Chapter 6 centers on disability design for blind students and the case of a Deaf tech employee. I found this the least compelling of the chapters, probably because this subject is within my domain of expertise. The call to action also seemed to be weaker than in other chapters; it concerned incorporating ethics into computer science, as well as the wins of *design thinking* (an approach to problem solving using a conceptual, user-centered approach), which are less lauded by some critics. Design thinking's emphasis on empathy and ethnography can be a problem for disabled people, who are often

seen by nondisabled people as pitiable, and designers would do better to talk to actual disabled people, which is something Broussard would recommend. She ends this chapter writing about the need to center BIPOC (Black, Indigenous, and People of Color) disabled people in these conversations, as those with multiple intersecting marginalized identities are often forgotten in tech conversations.

Broussard's other chapters talk about the gender biases encoded into systems, medical diagnostics that use race, and the process of making algorithms. Broussard is best when drawing from her experiences in data journalism and from her own life to illuminate the cases she shares. For instance, in the Introduction she explains fairness through the lens of childhood negotiation and bargaining; and in Chapter 8, she shares her experience of having an AI scan at the end of her breast cancer treatment and then later, running one of the many algorithms for breast cancer detection.

I almost skipped the chapter on cancer. As someone who has had cancer three times and gets scanned often, I sometimes skip things about cancer because it's too much. But this chapter, Chapter 9, was my favorite. Rather than being about the experience of cancer or its treatment, it is about how one *experiences* technology, and also about trying to understand how relevant algorithmic technologies work in this context. In this chapter, Broussard shows the weird quirks of current tech—at one point buying a CD-reader for her computer because her scans won't download off of her health portal in a readable way. She then tried out the algorithms for herself with her own scans, post-cancer treatment. The algorithm she tries (which she found out later was created by one of her neighbors) ended up with the "correct" result, confirming her breast cancer. However, the experience raised questions of access, compatibility, medical records systems, user interfaces, and the importance of human involvement in the process.

The penultimate chapter, Chapter 10 gives us a guide forward, describing algorithmic auditing as a powerful tool to help mitigate and manage currently unjust data systems. In the intro, she told us that:

> Social fairness and mathematical fairness are different.
>
> Computers can only calculate mathematical fairness.
>
> This difference explains why we have so many problems when we try to use computers to judge and mediate social decisions.

In order to address social fairness—and Broussard believes everyone should—we need to be looking for the biases that exist in the tech we've created. She writes about her own work as a consultant in algorithmic auditing, and shares how this approach is not a one-size-fits-all approach, which is why we need auditing suited to the needs of different projects and technologies. She highlights different watchdog and data audit projects that are doing good work in addressing data justice: Data for Black Lives, AI for the People, the Stop LAPD Spying Coalition, and more.

Social fairness and mathematical fairness are different. Computers can only calculate mathematical fairness.

She also highlights different regulatory changes and policies that are relevant to this work, including proposed legislation from the European Union from 2021 that calls for high-risk AI to be controlled and regulated, and that would divide AI into high-risk and low-risk categories for the purposes of monitoring and regulating those using high-risk AI. The proposal asks for a ban on facial recognition for surveillance, criminal justice algorithms, and test-taking software that uses facial recognition, which coincides nicely with the Ban the Scan proposal from Amnesty International.

The book is not a historically oriented work on bias in computing or in medicine. Nearly all of Broussard's cases and material come from the past decade, if not the past few years. She draws from other work that speaks to similar themes, such as that of Mar Hicks, Safiya Noble, and Ruha Benjamin. What this book offers is an orienting view that is both broad and specific on algorithmic biases and informed solutions for how to grapple with where technology is right now.

Broussard's action items emerge over each chapter and don't converge on just one approach. We need to be thinking about a variety of approaches, including outright bans of and policymaking on some types of algorithms due to their potential for great harm, as well as a willingness to slow down or abandon the adoption of algorithms in educational contexts. Other approaches to addressing algorithmic bias involve automated auditing, as long as such programs are "updated, staffed, and tended" using software packages such as AI Fairness 360 or platforms such as Parity and Aequita, which have mitigation algorithms designed to de-bias; Broussard herself works with O'Neil Risk Consulting and Algorithmic Auditing (ORCAA) on a system called Pilot for automated auditing. ORCAA advocates for internal audits and has worked with Joy Buolamwini's Algorithmic Justice League to address algorithm analysis with an intersectional framework, looking at different subgroups to assess performance and to identify where an algorithm can produce harm.

External audits are another approach; these are audits conducted by external agencies or watchdogs, as well as investigations triggered by whistleblowers. Exposing.ai, one watchdog group, lets you find out if your photos are being used to train facial recognition. Some external audit groups, such as The Markup, fact-check tech firms and demonstrate an important role for data journalists and investigative journalism in this area. Broussard also applauds Ruha Benjamin's idea of tech discrimination as an "animating principle" that makes it possible to see where technology may harm people or violate their rights. In all of these approaches, we need work that is aware of unchecked algorithms generating unfairness and bias—and a wariness of technochauvinists who mistakenly think they've produced a neutral tool.

More than a Glitch speaks to a wide audience, while also nodding to important scholarship if readers seek more depth. Broussard keeps the book short for all that she covers, which makes it a great introduction and reference. Her description of technochauvinism is not presented as a theoretical worry, but as a very real problem, made clear in case after case she references without excess theory building; for that, she points to other scholarly literature in the field that people can read. Thus, *More than a Glitch* is a useful companion to other literature, and it is also useful by itself as an introductory text. The cases are true and startling, in a way that invites conversations and reflections that we need to be having.

Ashley Shew is an associate professor of science, technology, and society at Virginia Tech. She's interested in the stories we tell ourselves about technology, and where we get the stories wrong.

Photo by C. Vernon

Erica Thompson is a mathematical modeling expert and author of *Escape from Model Land: How Mathematical Models Can Lead Us Astray and What We Can Do About It* (Basic Books, 2022).

Models and Mathematics: Q&A with Erica Thompson

Jaime R. Herndon

Erica Thompson's December 2022 book, Escape from Model Land: How Mathematical Models Can Lead Us Astray and What We Can Do About It, *explores the limits of mathematical models, and how we can use them in smarter, better ways. She is associate professor of modeling for decision making at University College London's Department of Science, Technology, Engineering, and Public Policy. She's also a fellow of the London Mathematical Laboratory and a visiting senior fellow at the London School of Economics' Data Science Institute. Thompson spoke with book review editor Jaime Herndon.*

Models: That term can mean a lot of things, especially for someone outside of the field. What kinds of models do you work with?

Well, not the glamorous kind that you'd see on a catwalk! I'm interested in scientific and mathematical models, simulations, or representations. That includes very complex physical models like climate and weather models, but also much simpler models like the basic infection models in epidemiology and even conceptual nonquantitative models like "If the price of a product goes up, then fewer people will want to buy it." Models are used to help us think about a subject, to make predictions of how a system might change over time, and to try to work through what the consequences would be if we intervened in some way. In science, we rely on models, but they are also increasingly important in policymaking, business decisions, and so on. The more data we have available, along with the computing power to collect and analyze that data, the more we tend to construct and use models.

You look at how we use models to make real-life, real-world decisions. Can you talk about why this is so important?

From the very brief examples I just gave, you can see the range of decision-making that is supported by models. The weather forecast on the phone in your pocket is produced by a very complex model, and you might use it for all sorts of everyday decisions; emergency service organizations probably use it for much more consequential decisions. Climate models are used to support decision-making about infrastructure investment, emissions policies, and international coop-

eration. Epidemiological models, from the very simple to the hugely complex, have supported public and private decision-making at all stages of the COVID-19 pandemic. And models are increasingly forming the bedrock of business analytics and forecasting, including the rapid use of new machine-learning techniques to create different kinds of models.

You coined the term the hawkmoth effect, which has to do with "the sensitivity to structural model formulation." How does this relate to models and decision-making, and what does this mean for real-world decisions? What are some examples of the hawkmoth effect in action?

Most people have heard of the butterfly effect, which is the idea that complex systems can be very sensitive to changes in the initial condition ("the flap of a butterfly's wings in Brazil can trigger a tornado in Texas"). From the point of view of forecasting, this means that if we get the initial conditions slightly wrong, then the forecast could become unreliable on some timescale. The hawkmoth effect is a similar idea, but relating to the accuracy of the model rather than the data. If the model is slightly "wrong," even by just a tiny amount, then the forecast it makes could be significantly wrong if you are predicting far enough into the future.

Of course, the problem is that just as all data are subject to uncertainties (measurement errors), also "all models are wrong" in the sense that we can never really know that we have fully represented absolutely everything that is important about the system. In some cases, like the weather forecast, we have lots of past evidence from our successful predictions, which help us have confidence; we know that tomorrow's weather forecast is pretty good, next week's is indicative but not all that reliable, and we wouldn't even look at the forecast six months in advance because we know, based on past evidence, that it wouldn't be of any use. So, for the weather forecast, we have a good idea of when it is useful and when it isn't, but we don't necessarily have such a clear idea of the limitations of other models and other kinds of predictions about the future.

In what ways are models helpful, and in what ways have models been not so helpful, or even counterproductive, in real life? What can we learn from these examples?

Models are incredibly helpful. Models can send us to the Moon and they power the modern world, from your electricity supply to your social media feed. And I'd go further to say that they are very much part of the way that we think. Both at the mathematical end of institutional and organizational decision-making, where you might think of weather models, economic models, pandemic models, or business models, but also at the conceptual level. I argue in my book that models are primarily useful as metaphors and aids to thinking, rather than prediction engines, and so you might think of the "flatten the curve" concept that was so important in the spring of 2020, or more qualitative concepts like the idea that the national budget either is or is not like a household budget. In that light, I hope it becomes clearer how models also change the way that we think: If you have a particular model (mathematical or conceptual) for something in the real world, then you use that model as a tool for understanding, and as a tool for communicating with other people. It can be illuminating by helping us to think in new ways, but the model and the limits of the metaphor also constrain the ways that we are able to think about the system.

Models are primarily useful as metaphors and aids to thinking, rather than prediction engines.

To take an example, if you construct a simple epidemiological model for a disease outbreak, you are focusing (probably rightly) on the infection and its consequences, and you might choose to model people as statistically representative populations or as interacting individuals with different characteristics, and so on. You might have schools and hospitals and prisons in your model, or you might not. Now if a politician comes and asks you what can be done about the outbreak, you will frame your advice in terms of the information you have from your model and the kinds of interventions that can be represented in the model. This is fine, and completely reasonable, as long as the decision-maker also has access to other kinds of sources of information about the impacts of different policies. Epidemiology is a very politicized example, which perhaps makes the shortcomings of a model like that more obvious. The models from the COVID-19 pandemic weren't counterproductive, they were able to contribute to decision-making. But what we can learn from the pandemic is that for highly contested and complex decisions, we need a more diverse range of models, and more effective ways to think about the insights gained from models.

What do you think is the most important thing we should keep in mind for working with models in the future, and where do you see your work going?

The most important thing to keep in mind when working with models is that any model can only give us one perspective. It can't tell the whole story. A photograph is great to show you what someone looks like, but it doesn't tell you their political opinions or what they want to have for lunch. When we make and use models, we need to keep in mind what they are good at and what they aren't good at.

My own work has two strands, a mathematical strand about the statistics of model calibration and evaluation, and a more sociopolitical strand about the value judgements that are embedded in models—both illustrated through case studies of different kinds of models, from public health to finance and climate change. I'm working on bringing these two strands closer together, to learn from each other and hopefully to improve the usefulness of the modeling methods that are so important for today's decision-making.

Jaime Herndon is American Scientist*'s acting book review editor.*

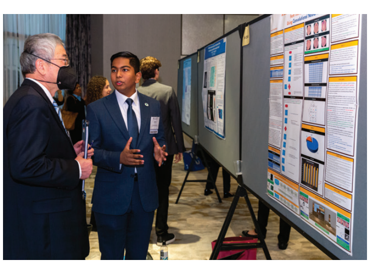
Sigma Xi Today
A NEWSLETTER OF SIGMA XI, THE SCIENTIFIC RESEARCH HONOR SOCIETY

Student Research at IFoRE '23

Calling all graduate, undergraduate, and high school students—present your research and win awards at the 2023 International Forum on Research Excellence (IFoRE) in Long Beach, California!

Sigma Xi is currently accepting student research submissions for oral and poster presentations across all STEM disciplines. Show us your research! When you enter the student competition at IFoRE, you'll get to showcase your work to some of the world's most distinguished scientists, engineers, academics, and scientific professionals in your field. Gain the invaluable experience of communicating your research to a prestigious panel of judges, receiving essential mentoring feedback, and competing for the prestigious IFoRE awards given out in all student categories. Most student awards will also include monetary prizes and Sigma Xi membership.

For more information, please visit **experienceIFoRE.org/submissions**.



Sigma Xi Today is managed by Jason Papagan and designed by Chao Hui Tu.

From the President

Sigma Xi Scientists: Building the Future upon Scientific and Ethical Foundations

We are scientists. Every year, our mission becomes more complex. We must confront global warming; the depletion of underground reserves of clean water; the evolution of viruses and the resistance of bacteria to antibiotics; interrupted supply chains; and inadequate resource management. We are surrounded by, and we are ourselves, struggling humans on whom far too many demands have been placed, some of us with children who are expected to be perfect even though they are constantly measured by impossible standards.

Humans started using science to improve our living conditions at least 4,000 years before the fathers of the Enlightenment codified the basic steps of the scientific method. In ancient Mesopotamia, the priestess Princess Enheduanna charted the paths of heavenly bodies so she could determine the best times to plant and harvest crops. There is a direct line from her important work to the great challenges we face today.

Vision is central to the work of scientists. Having a sense of vision is not dependent upon the eyes, but rather, the brain and its ability to extract knowledge from experience. Through interaction with our environment, we acquire and selectively store information that allows us to create mental images. When observation inspires visualization, we can design plans to surmount challenges and achieve goals toward creating a better world.

Another central facet of scientific work is communication. Since before the times of Enheduanna, humans have developed complex, flexible communication systems so that humans may partake of available knowledge. Thus, the vision of one generation can be passed to the next.

For the scientific enterprise to prosper in our complex modern world, we must devise and implement teaching strategies that will allow young people to develop autonomous skills to evaluate the quality of information. They must learn to discriminate what information is useful to collect, store, and incorporate into their working knowledge.

Isaac Newton famously said that he owed his greatness to all the great people who preceded him and on whose shoulders he stood. He was referring to the skills of scientific vision and knowledge acquisition, under the constraints of moral principles, and to the incorporation of the available body of knowledge to the task of improving the human condition for future generations. The scientific method of addressing and, with great effort, solving problems requires primarily the ability to think. We must strengthen the capability of students at all levels of the educational system to develop independent judgment, moderated with the humanistic goal of improving the human condition everywhere. The members of Sigma Xi are just the community to chart the way.

Marija Strojnik

Marija Strojnik

Sigma Xi Welcomes Adriana Bankston as New Senior Fellow, Civic Science and Public Policy

Sigma Xi, The Scientific Research Honor Society, in partnership with the Rita Allen Foundation, welcomes Adriana Bankston in the position of senior fellow, civic science and public policy. In this new role, Bankston will lead a project that examines science policy engagement at the state level and that determines the skills, knowledge, and resources required by scientists to successfully influence public policy.

The Civic Science Fellowship program seeks to broaden engagement with science and research to inform policy and develop solutions to societal challenges. As senior fellow, Bankston will be connected to a national network of fellows from diverse backgrounds working on a variety of multidisciplinary projects. She will also work closely with Sigma Xi leadership to develop, launch, and manage an online platform to connect current policy players, showcase policy-engaged organizations, and empower individuals from diverse backgrounds to successfully engage and achieve societal change through policy impact at the state level.

Since 2021, Bankston has served as CEO and managing publisher of the *Journal of Science Policy and Governance*, an internationally recognized nonprofit organization and peer-reviewed publication dedicated to empowering early career scientists, engineers, and policy professionals in international science policy debate. She is also a biomedical workforce and policy research investigator at the STEM Advocacy Institute, where she works to cultivate the next generation workforce through science policy. Finally, Bankston is a fellow with Advancing Research Impact in Society (ARIS), where she received the inaugural ARIS Emerging Broader Impacts Leader Award in 2022. She holds a PhD in biochemistry, cell biology, and developmental biology from Emory University.

"I am honored to work with Sigma Xi in this new role and contribute to enhancing the connection between science and society through public policy," Bankston said. "Local engagement of scientists with the policymaking process is critical to developing the next generation of leaders in the field and empowering them to improve society through policy change."

"We are thrilled to welcome Dr. Bankston to the Sigma Xi team," said Jamie Vernon, executive director and CEO of Sigma Xi. "Her expertise will advance Sigma Xi's efforts to identify synergies within the science policy training ecosystem, and increase efficiency and capacity for creating evidence-informed policy at the state level."

2023 Student Research Showcase Winners

Sigma Xi, The Scientific Research Honor Society presented awards for its Student Research Showcase on May 22, 2023. The virtual competition included 212 student presentations across 12 disciplinary categories. Monetary awards were given for first and second place winners in the high school, undergraduate, and graduate divisions. Additionally, a people's choice award was given to the presentation with the most likes on the competition website.

The Student Research Showcase is an annual virtual competition aimed at building students' science communication skills so they can convey the value of their research to technical and nontechnical audiences. Participants submitted abstracts for entry into the competition in early spring. During a month-long evaluation period, students built websites, videos, and slideshows to present their research to a panel of judges and public audiences. Judges' evaluations were based on how well the students communicated enthusiasm for their projects; explained the significance of their research; used text, charts, and diagrams; and responded to questions. Congratulations to all 2023 participants and winners!

High School Division

First Place (4-way tie)

Cate DeVane
Engineering
"Hydrogel Material Characterization for Targeted Drug-Delivery Technologies"

Calvin Matthew
Engineering
"3D Printing Personalized Knee Implants: Novel Biocomputational Models for Meniscus Tear Regeneration"

Kaitlyn Wang
Physics & Astronomy
"ExoScout: Discovery of the Smallest Ultra-Short-Period Exoplanets"

Christopher Chen & Minghao Zou
Physics & Astronomy
"Detecting Faint, Fast-Moving Near-Earth Objects"

Second Place

Likhitha Selvan
American Heritage School, Plantation, Florida
Physiology and Immunology
"Unraveling the Mysteries of Inflammasomes in Alzheimer's Disease"

Undergraduate Division

First Place

Joseph Benevento
Environmental Sciences
"Abundance and Diversity of Oyster Microbiomes"

Second Place

Matthew Franolich
Environmental Sciences
"Oyster Pathogen Monitoring Using Third Generation Sequencers"

Graduate Division

First Place

Vana Mahabir
Physiology and Immunology
"Porcine Model of Transgender Men and PCOS"

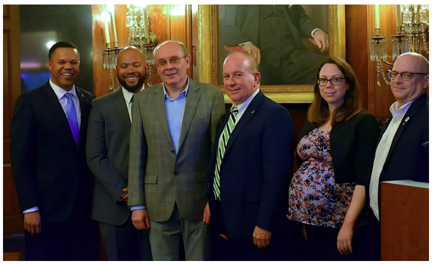
Second Place

Tracy Dubin
Human Behavioral and Social Sciences
"Repetitive Transcranial Magnetic Stimulation (rTMS) Impacting Working Memory"

People's Choice Award

Nobuhiro Komatsuda
Human Behavioral and Social Sciences
"School-Related Stressors and Psychosomatic Symptoms"

Pariser Global Lectureship for Innovation in Physical Sciences



On April 27, 2023, Sigma Xi held the second annual Pariser Global Lectureship event at the Morehead Planetarium and Science Center on the campus of the University of North Carolina at Chapel Hill (UNC).

The lectureship reflects the legacy of pioneering chemist Dr. Rudolph Pariser and the creativity and research excellence exemplified throughout his life and career. Designed to connect chapters and members with thought leaders in science and engineering, the Pariser Global Lectureship recognizes researchers worldwide for their capacity to bridge the gap between basic and applied research for the betterment of humanity.

Hosted by the UNC chapter of Sigma Xi, the 2023 event treated members and guests to an evening of networking, hors d'oeuvres, chapter recognition, and compelling presentations. The evening's featured speaker was Dr. Christopher Clemens, who serves as provost and Jaroslav Folda Distinguished Professor of Physics and Astronomy at UNC. Dr. Clemens shared insight on his research with spectroscopy tools that measure the composition of rocks and other debris in planetary systems beyond our own.

Sigma Xi chapters worldwide are invited to serve as the host of future Pariser lectureships funded by the program's endowment. Visit **sigmaxi.org/pariser** to learn more.

Sigma Xi Installs New Chapter at Elmezzi Graduate School of Molecular Medicine

On March 30, 2023, Sigma Xi installed its newest chapter at the Elmezzi Graduate School of Molecular Medicine at Northwell Health. The ceremony was held in person in Manhasset, New York, and presided over by Sigma Xi then President-Elect Marija Strojnik. It was a celebration of the new chapter's officers, members, and commitment to growth and advancement of the school's research enterprise.

Annette Lee, dean of the Elmezzi Graduate School, delivered the inaugural presidential address. She will lead a group of founding members that include Yousef Al-Abed, Lance Becker, Betty Diamond, Daniel Grande, Christine Metz, Barbara Sherry, Bettie Steinberg, Kevin Tracey, and Ping Wang. The founding members will guide the chapter through its initial years of development and activity, including recurring meetings, events, membership growth, and participation in Sigma Xi's annual conference, the International Forum on Research Excellence (IFoRE).

The ceremony was also attended by Sigma Xi's executive director and CEO, Jamie Vernon, and Sigma Xi's director of membership and chapters, Richard Watkins. The CEO of Northwell Health, Michael Dowling, delivered the keynote address.

The Elmezzi Graduate School of Molecular Medicine at Northwell Health, located in Manhasset, New York, is an individually tailored and accelerated three-year doctoral program awarding a PhD degree to individuals who hold an MD or equivalent. The graduate school directly addresses the crucial need for physician-scientists who are trained to conduct translational and clinically relevant research.

FACES of GIAR

April Stabbins

Grant: $1,000 in Fall 2020
Education level at time of the grant: PhD student

Project Description: The project sought to investigate how certain coral species interact with methane seeps in the deep sea. By receiving this grant, I was able to collaborate with researchers from other laboratories and complete more detailed analysis with advanced microanalytical techniques that otherwise would not have been possible.

How did the grant process or the project itself influence you as a scientist/researcher? My first application was not accepted, but this outcome caused me to go back and review my proposal in detail to make adjustments. Being able to take a step back and review how I describe my work to others, especially those not in my field, has influenced how I have applied for other grants and how I write manuscripts.

Where are you now? I am a fifth year PhD candidate at Temple University in Philadelphia. I plan to graduate in the next year and am now looking for a postdoctoral position.

Chhandak Basu

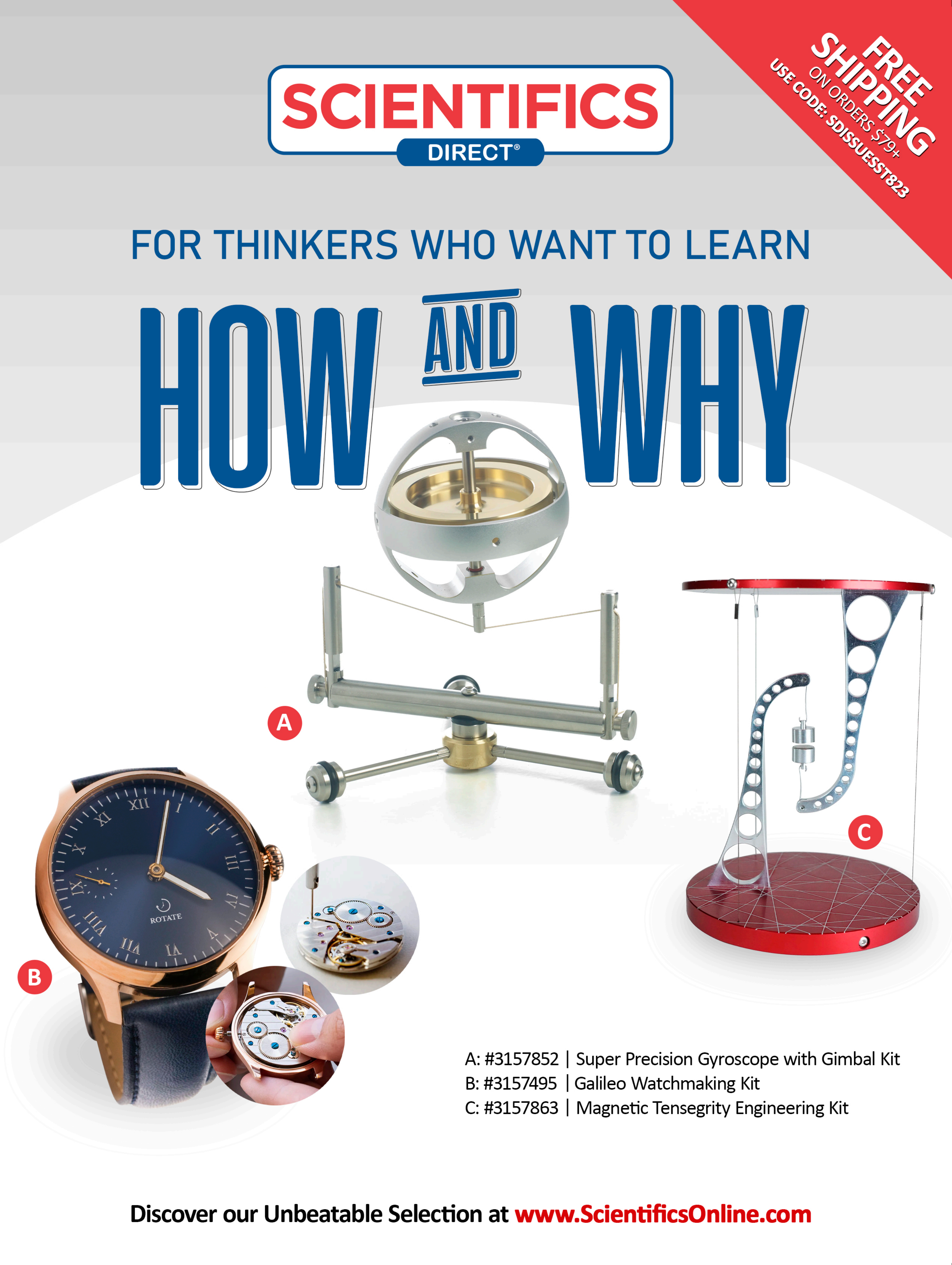
Grant: $1,000 in Fall 2002
Education level at time of the grant: PhD student

Project Description: Thanks to the GIAR funding, I was able to visit and work in the laboratory of Brian Friskensky, Department of Plant Science, University of Manitoba, Canada. As part of my PhD thesis, I evaluated several gene promoters that are suitable for the genetic transformation of plants. I was fascinated to examine gene expression in plant cells at a transcriptome level. However, this was not a direct topic of my research. My major professor was very supportive of my idea to visit another laboratory to learn techniques for studying stress-responsive gene expression in plants.

As a result, I visited Dr. Friskensky's lab to learn how transcription analysis with cDNA-macroarray can be used to study gene expression patterns after pathogen attack in canola cells. My research involved preparing pathogen-induced cDNA from canola cells and mastering macroarray techniques. Various genes were nonradioactively labeled and hybridized onto membranes containing cDNA from canola cells. In addition to understanding pathogen-induced gene expression in plants, the genotyping data could also be used to develop disease-resistant crops.

How did the grant process or the project itself influence you as a scientist/researcher? Working in a different lab abroad, outside my familiar environment, was an experience I will never forget. Having this research experience motivated me to become an independent researcher. Within five years of receiving the GIAR fellowship, I attended a week-long workshop at the University of Arizona, Tucson, to learn microarray. As of now, my lab focuses on transcriptomic-level gene expression studies in plants, and therefore I am indebted to Sigma Xi for granting me the GIAR award.

Where are you now? I am currently a professor of biology at California State University, Northridge. As part of my research, I study how plant cells respond to environmental stress at the molecular level. Additionally, I have a research affiliate appointment at NASA's Jet Propulsion Laboratory. I conduct research at NASA on how microbes survive in hostile environments.


SCIENTIFICS
DIRECT®
FREE SHIPPING
ON ORDERS $79+
USE CODE: SDISSUESST823
FOR THINKERS WHO WANT TO LEARN
HOW AND WHY
ROTATE
A
B
C
A: #3157852 | Super Precision Gyroscope with Gimbal Kit
B: #3157495 | Galileo Watchmaking Kit
C: #3157863 | Magnetic Tensegrity Engineering Kit
Discover our Unbeatable Selection at www.ScientificsOnline.com